DOCUMENT RESUME

ED 324 364 TM 015 631

AUTHOR	Baghi, Heibatollah; Ferrara, Steven
TITLE	A Comparison of IRT, Delta Plot, and Mantel-Haenszel Techniques for Detecting Differential Item Functioning Across Subpopulations in the Maryland Test of Citizenship Skills.
PUB DATE	22 Mar 89
NOTE	69p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE	Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	Black Students; Citizenship Education; Comparative Analysis; Comparative Testing; Grade 9; *High School Freshmen; High Schools; *Item Bias; *Item Response Theory; *Multiple Choice Tests; *Racial Differences; Sample Size; *Sex Differences; State Programs; Testing Programs; White Students
IDENTIFIERS	Delta Plot Method; Mantel Haenszel Procedure; *Maryland Test of Citizenship Skills; Rasch Model

ABSTRACT

Use of item response theory (IRT), the delta plot method, and Mantel-Haenszel techniques to assess differential item functioning (DIF) across racial and gender groups associated with the Maryland Test of Citizenship Skills (MTCS) is described. The objective of this research was to determine the: effect of sample size on results from these three DIF techniques; degree of relationship among these DIF statistics; and degree to which they identify the same items as biased. The data for the study include item responses from one form of the 1988 edition of the MTCS. The MTCS consists of 45 multiple-choice items that assess students' knowledge and skills in 3 domains: constitutional government; politics and political behavior; and principles, rights, and responsibilities. The MTCS was administered to 50,000 ninth graders during January and February of 1988. The analyses were performed on representative samples of 1,000, 750, 500, and 200 first-time test takers. It is concluded that no MTCS items are functioning differentially in either black/white or male/female comparisons. Plots of item difficulty estimates for black/white and male/female comparisons show nearly perfect linear relationships in both groups. Agreement, as indicated by rank order correlations across DIF techniques, is very high between Rasch and Delta Plot DIF indices for all sample sizes in both black/white and male/female comparisons. In terms of agreement regarding biased and unbiased items, agreement with the three-parameter DIF index is highest for the Delta Plot and Rasch techniques. A 30-item list of references, 19 data tables, and 30 figures are included. (TJH)

A Comparison of IRT, Delta Plot, and Mantel-Haenszel Techniques for Detecting Differential Item Functioning Across Subpopulations in the Maryland Test of Citizenship Skills

Heibatollah Baghi
Steven Ferrara

Maryland State Department of Education
March 22, 1989

Our thanks to Jerry Cunningham and Sandy Gedeik for their skill and diligence in designing, writing, and revising the programming necessary for these analyses

Paper presented at the annual meeting of the American Educational Research Association, March 28, 1989, San Francisco

The study of test items that function differently for subpopulations of examinees is of concern to test developers. This concern is especially critical in competency-based testing, where graduation certification is contingent on passing one or more tests. Differential item functioning (DIF) was originally called item bias. Many researchers have attempted to define it clearly. In educational measurement, the term bias is used in reference to <u>tests</u> and their use usually for selection and classification, or individual <u>items</u> and their effect on total test scores. Test items may be considered biased when a minority group scores disproportionately lower than a reference group due to factors other than ability. Cleary and Hilton (1968) defined item bias as an interaction between item and group in terms of analysis of variance procedures. Angoff and Ford (1973) considered an item biased if the item difficulty index is significantly higher or lower in one group than in another group. Scheuneman (1979) considered an item biased if, for all examinees having the same score on a test that includes that item, the proportion of examinees answering the item correctly is substantially different for various subpopulations being considered.

Clearly, the definition of item bias is dependent in part upon the techniques that are used to find differentially performing items. For example, when using item response theory to investigate item bias, an item is considered unbiased if the item characteristic curves (ICCs) for the item are the same for both subpopulations (Crocker & Algina, 1986, p. 377). In that case, among individuals with the same ability score, the items are equally difficult for members of both subpopulations. Somewhat similarly, in chi-square techniques, an item is considered unbiased if within a group of individuals with total test scores in the same test score interval, the proportion of individuals responding correctly to the item is the same for subpopulations (Crocker & Algina, 1986, p. 383). Transformed item difficulty techniques (e.g., Delta Plot) base the definition of DIF on the notion that, when items are ranked according to difficulty, unbiased items will be ordered the same in two compared groups. The assumption here is that bias is indicated by a significant group difference in the relative difficulty of an item rather than by a large group difference in item difficulty means and standard deviations (Osterlind, 1987, p. 28). A widely accepted definition for DIF is that an item is considered unbiased if examinees with equal ability, but from different subpopulations, have equal probability of responding correctly to the item (Angoff, 1982).

A variety of techniques for detecting DIF have been proposed in the literature. Hills (1977, 1981, 1982) identified more than 40 techniques for this purpose and grouped them into nine general types: (1) methods for comparing plots of transformed item difficulties; (2) analysis of variance methods; (3) chi-square methods; (4) foil methods, which involve examining the differential response patterns of various groups of examinees to item foils in order to find alternatives which overly attract or repel a particular group; (5) correlation methods, which involve comparison of the reliabilities of a test when the reliability is estimated for each group separately; (6) item response theory methods; (7) factor analysis methods; (8) methods based on experimental manipulations; and (9) construction methods to ensure unbiased tests.

These techniques are different but are concerned with similar concepts of bias. They often produce different results because of theoretical and practical reasons. Thus, many studies of DIF techniques in the past several years have been devoted to comparing different techniques. The numerous techniques proposed for detecting DIF have been narrowed down in recent years to several of the most promising. There exist several comprehensive reviews of the DIF literature (Burrill, 1982; Ironson, 1982; Rudner, Getson, & Knight, 1980; Osterlind, 1987; and Shepard, Camilli, & Williams, 1985). The consensus from this research is that "the ICC approach is the most generally valid of all biased item detection methods" (Osterlind, 1987, p 69). Item response theory (IRT) techniques are the theoretically preferred procedures for detecting DIF because they least confound real mean differences in group performance with bias (Lord, 1977). The sample invariance property of IRT provides a theoretical framework of how DIF is defined and detected in a test. ICCs describe the relationship between item difficulty and examinee ability in terms of the probability of responding correctly. If an item has the same meaning in two comparison groups, then the probability of a correct response should be the same for examinees of equal ability from different groups. Although the IRT approach is superior theoretically, there are practical problems in using it. For example, IRT computer programs are costly and complicated to use. In addition, the three-parameter model requires a minimum of 1,000 cases per group (i.e., for LOGIST) to estimate item parameters, a requirement that often is difficult to meet in minority samples. As a result, other techniques that are not limited by difficult sample size requirements have been developed; for example, chi-square techniques that are considered approximations to item response theory techniques (Scheuneman, 1979; Holland & Thayer, 1986). An advantage of chi-square techniques is that they are easier to apply than IRT techniques and do not require large sample sizes. However, the relationship between the size of the Mantel-

Haenszel chi-square and sample size has not received attention beyond mere mention (see, for example, Raju, Bode, & Larsen, 1989, p. 12). DIF procedures that are clearly recommended in the literature are IRT methods, Mantel-Haenszel chi-square techniques, and Angoff's delta plot method (Shepard et al., 1985, p. 84).

In this paper we describe briefly and compare three techniques for detecting DIF: item response theory (IRT, using the three parameter model and the Rasch model), Delta Plot, and Mantel-Haenszel (MH) chi-square techniques. We compare IRT and Mantel-Haenszel approaches because they are reputed to produce similar results (see Rudner, Getson, & Knight, 1980). We include the delta plot technique because it has been recommended as an alternative in situations where sample size or other practical considerations preclude the use of IRT or chi-square methods (Subkoviak, Mack, Ironson, & Craig, 1984). The objective of this research is to determine the (1) effect of sample size on results from these three DIF techniques, and (2) degree of relationship between these DIF statistics, and (3) degree to which they identify the same items as biased.

Method and Procedures

Data Source

The data for this study are item responses from one form from the 1988 edition of the Maryland Test of Citizenship Skills (MTCS). The MTCS consists of 45 multiple-choice items that assess students' knowledge and skills in three domains: Constitutional Government; Principles, Rights, and Responsibilities; and Politics and Political Behavior. Annual forms of the MTCS are constructed by sampling items from a large bank that has been calibrated using the Rasch model. Students must pass the MTCS, along with three other minimum-competency tests, in order to receive a Maryland high school diploma. The Maryland Functional Testing Program (MFTP) uses two approaches for detecting DIF: judgmental reviews and statistical analysis. Before newly written items are field tested, experts in ethnic and sex bias review their language and the situations they pose for potential sources of bias. The Delta Plot technique is used as a post-administration check for differentially functioning items. Flagged items are examined for potential causes of DIF before they are included in a student's score on the MTCS, and later resubmitted for review by bias specialists.

Construction of samples. The MTCS was administered to approximately 50,000 9th grade students during January and February, 1988. The analyses are performed on representative samples of 1000, 750, 500, and 200 first-time test takers. Random comparison groups (referred to as "random 1" and "random 2") of each of the four sample sizes were created by randomly selecting cases from the entire pool. White-black and male-female comparison groups of each of the four sample sizes were created by randomly sampling cases from within race and sex strata.

A critical assumption made for DIF techniques is that the test under scrutiny is unidimensional; that is, that all items measure the same latent ability, skills, and so forth. Investigating the unidimensionality assumption is problematic because experts do not agree on appropriate methodology and criteria for testing this assumption. In this study, the recommendations of Reckase (1979) for determining unidimensionality of a test were used as follows: (1) In a factor analysis of test items, the first unrotated principal component should account for at least 20 percent of total test variance; (2) The eigenvalue for the first principal component should be large relative to the eigenvalue for the next largest component.

In the next section of this paper we describe procedures for detecting DIF using IRT, Delta Plot, and MH techniques. In subsequent sections we describe results from implementing these three techniques in MTCS items and draw conclusions about the stability and agreement of the results.

DIF Techniques, Methods of Analysis, and Procedures

IRT Techniques and Procedures

According to item response theory, item parameter estimates are invariant with regard to the group used in the estimation. If an item's parameter estimates are different for different groups, according to the theory, then the item must be measuring more than a unidimensional ability assumed by the model. Therefore, item parameters that vary across subgroups indicate DIF. In this study we use graphical analysis for descriptive purposes and differences between ICCs to detect differentially functioning items.

Graphical analysis. Graphical analysis involves plotting difficulty estimates for each item (and discrimination estimates in the three parameter model) for a focal group (i.e., blacks, females) versus a reference group (i.e., whites, males). Item difficulty and discrimination estimates for comparison groups (blacks vs. whites and males vs. females) were plotted separately. This graphical analysis is recommended by Hambleton (1982) as a simple method to detect potentially biased items. Theoretically, if the item is functioning the same in both groups the difficulties in both groups should be identical, except for estimation and sampling error, and plotted points should tightly hug a best-fitting line.

Differences between ICCs. We examine differences between ICCs from the three parameter model and the Rasch model separately. Examining three parameter ICC differences involves six steps. First, item parameters for all items are estimated for two random groups using the PC-BILOG program (Mislevy & Bock, 1986). DIF results from analysis of these parameters provides a criterion for distinguishing real DIF, which may be caused by some form of bias against a subgroup, and apparent DIF due to sampling error. Second, item parameters were estimated for all items separately for each reference and focal group. Third, item parameters for the reference and focal groups were linearly transformed to the same scale. Fourth, using the item parameter estimates for the two random groups in step one as input, difference was calculated between the two ICCs for each item. Fifth, the absolute difference between the ICCs for each items and the mean, standard deviation, and 99 percent confidence interval for these absolute differences were found. Finally, confidence intervals were used as a baseline to identify extreme differences in the ICCs found with the majority and minority groups: any difference not contained in the confidence interval was considered an indication of DIF.

The BICAL computer program (Wright, Mead, & Bell, 1979) was used to estimate Rasch model icem difficulty parameters. Rasch model ICCs were also compared following the six steps described above. However, since Rasch model ICCs are parallel, the area between the ICCs for the same item in two groups is equal to th difference in the item difficulties (see Phillips & Mehrens, 1988).

Delta Plot Technique and Procedures

This Delta Plot technique introduced by Angoff (1972), is based on an item-by-group interaction as a measure of DIF. This method can produce spurious evidence of bias unless all items are equal in discrimination or the groups being compared do not differ in average performance. To solve this problem Angoff (1982) proposed modifications to correct for this source of error. We implement Angoff's modifications in this study by performing Delta Plot analyses on groups matched on total test score. Item p-values were computed separately for matched white-black and male-female groups. Item p-values were then converted to an interval scale by determining the normal deviate z-value associated with the p-value and transformed to delta values with mean 13 standard deviation 4. Delta values for each pair of comparison groups were plotted for all items. Paired dalta values falling some critical distance from the plot's principal axis may be regarded as contributing to item-by-group interaction (Angoff, 1982). The perpendicular distance of the paired Delta values from the principal axis line is the bias index. In this study, items more than ± 1.5 z-score units from the fixed line are considered to be functioning differentially (see Rudner, 1977).

Mantel-Haenszel Technique and Procedures

The Mantel-Haenszel statistic (MH; see Holland and Thayer, 1986) is based upon two-by-two contingency tables for calculated for several total test score categories. This statistic is distributed as a chi-square with one degree of freedom and is considered by some to be the most powerful unbiased test of DIF (Cox, 1970). For the MH technique, a computer program uses scored item responses from reference and focal groups as input. The program calculates for each item a: (1) Mantel-Haenszel chi-square statistic (see Holland & Thayer, 1986, p. 8), and (2) difference measure called the common odds ratio across two-by-two tables (see Holland & Thayer, 1988, p. 134). If the MH chi-square statistic is significant, the item is considered to be performing differentially for one of the compared groups. In addition, if the difference measure is greater than one, the item is performing differentially in favor of the reference group; if it is less than one, then it is performing differentially in favor of the focal group. The present research uses five test score intervals in calculating the MH indices. Rudner, Getson, and Knight (1980) found that MH techniques using five intervals were as effective as IRT methods under most conditions.

Procedures

The three parameter IRT procedure was implemented in the samples of 1000 examinees. Rasch, Delta Plot, and MH procedures were implemented in the samples of 200, 500, 750, and 1000 examinees to investigate the effects of sample size on results. Finally, reference-focal group comparisons were made for blacks versus whites and males versus females. The different DIF techniques were evaluated in terms of stability, agreement, and practical and other limitations as described below.

In this study we compare the IRT, Delta Plot and Mantel-Haenszel techniques in terms of their (1) <u>stability</u> (concordance of each DIF method with itself across different sample sizes), and (2) <u>agreement</u> (concordance of DIF methods with results from the three parameter DIF approach and with one another). We evaluate concordance by examining (1) correlations between DIF indices, and (2) proportions of items identified by pairs of DIF techniques as biased and unbiased. We also evaluate these methods according to practical and other limitations (e.g., required sample sizes, availability of software).

In this study using real test data, there is no single criterion for identifying biased and unbiased items. However, unlike DIF studies that use simulated data, there is no means in this study for pre-identifying biased items. Instead, we identify biased items using the three parameter IRT technique decribed above and compare results from the other three methods to these results. Previous research (Shepard et al., 1985) indicates that using three parameter IRT techniques produces superior results in both real and simulated test data.

Results

Table 1 shows raw score means and standard deviations for all groups of examinees and all sample sizes. Item p-values, point biserial correlations, and test reliabilities are also presented for all groups and sample sizes. An examination of these means indicates that the white students scored higher on this test than black students across different sample sizes. However, the mean scores for male and female students are very similar. Internal consistency reliabilities are quite high and similar for the different groups and sample sizes.

Table 1 about here

The unidimensionality of the MTCS was tested by extracting principal components from item correlation matrices computed on randomly selected samples of 1000 students. The proportion of variance accounted for by the first principal component for the white group is 19 percent, for the black group 18 percent, for male group 20 percent, and for the female group equaled 19 percent. In all four analyses the eigenvalue of the first principal component was at least four times as large as the next largest component. Thus, the eigenvalue criterion for unidimensionality recommended by Reckase (1979) is easily met for the MTCS in all comparison groups used for this study, although the explained variance criterion is not.

Results from Each DIF Technique

Three-Parameter IRT DIF Technique

Graphical analysis. All analyses using the three parameter model are based on samples of 1,000 students. Table 2 presents item difficulty, discrimination, and guessing parameter estimates for three pairs of comparison groups: random groups 1 and 2, whites and blacks, and males and females. Plots of item difficulties and discriminations for these comparison groups appear in Figures 1-6. (Reference groups always appear on the Y-axis, focal groups on the X-axis). Correlations of item parameters for each pair of groups also appear in each plot. The graphical results suggest that no items on the MTCS are racially or sexually biased in terms of difficulty levels. In fact, the plots show nearly perfect linear relationships in the race and sex comparison group analyses. The correlation coefficients for discrimination estimates are not as high; the highest of these correlations are .85 in the random comparisons, .84 for black-white comparisons, and .68 for male-female comparisons.

Table 2 and Figures 1-6 about here

ICC differences. Differences between ICCs for pairs of groups was examined by comparing the area between the ICCs for the two independent random samples to the area between the ICCs for white-black and male-female samples. This method takes into account differences between compared groups in item difficulty, discrimination, and "guessing" as reflected in ICCs.

The confidence intervals for the absolute differences between the two random groups were used as a baseline to identify extreme differences in the ICCs found in black-white and male-female samples. Any differences not contained in these confidence intervals were considered an indication of a differentially functioning item. Means, standard deviations, and 99 percent confidence intervals for differences and their absolute values are reported in the first three columns of Table 3. Only three out of 45 items in the white-black comparison are outside of the confidence interval and identified as potentially biased against blacks. Two items were detected using this procedure as potentially biased against females. These five items are flagged for review by the bias committee judgmental review.

Table 3 about here

Rasch Model DIF Technique

Graphical analysis. To illustrate the Rasch DIF results graphically, Figures 7-10 depict plots of item difficulties for the random, race, and sex comparison groups (N=200). Figures 11-22 depict the similar plots for sample sizes of 500, 750, and 1000. The plots show nearly perfect linear relationship between the groups of examinees, in both sex and race analyses. Plots and correlation coefficients identify no items to be functioning differentially for race or sex subgroups in terms of difficulty level.

Figures 7-22 about here

ICC differences. The area between the two ICCs for each item for two independent random samples, white and black samples, and male and female samples across four different sample sizes were also examined. Again, a confidence interval for the absolute differences between two random groups was used as a baseline to identify extreme differences in the ICCs for the race and sex samples. Means, standard deviations, and 99 percent confidence intervals for differences and absolute differences are reported in columns 4-15 of Table 3. In the sample size of 1000 four items were identified as potentially biased against black students. In the sample size of 750 three deviant items were identified, and in the sample sizes of 500 and 200 only one item was identified as potentially biased against blacks (the same item). No items were identified as potentially biased in the sex group comparisons in the sample size of 1000. A single item was identified as biased against females in the samples 750, 500, and 200 examinees.

Delta-Plot Technique

Item delta plots for black-white and male-female samples matched on total score and samples of 200, 500, 750, and 1000 are shown in Figures 23-30. Item statistics and DIF indices for the various comparison groups across sample sizes are reported in Tables 4-11 which accompany the plots. The last column in each table contains deviations from a regression line, referred to as "Bias*." The critical value of this index for classifying an item as biased is greater than ± 1.5 z-score units from the line. No items on the test appear to be racially or sexually biased in the sample of 200. However, in the sample of 1000 students two items appear racially biased, and in the sample of 750 students four items appear racially biased. Two of those items were the same items found to be biased in the sample of 1000 students. Two other items appear to be racially biased in the sample of 500. Regarding potential sex bias, no items on the test appear to be biased in the samples of 200, 500, 750. However, in the sample of 1000 students, two items were identified.

Figures 23-30 and Tables 4-11 about here

Mantel-Haenszel Technique

The MH technique also is performed on representative sample of 1000, 750, 500, and 200 students. The output from the MH analysis includes a chi-square statistic and a difference measure. The chi-square statistic is compared to a chi-square with one degree of freedom. The difference measure indicates the direction of the bias. The results in black-white samples show that as the sample size used in an analysis was decreased, a pattern developed in the chi-square statistics; that is they became smaller and identified fewer items as biased. A similar pattern was not observed in male-female samples. These results in black-white samples show a dependence between the size of the chi-square statistic and the sample size used in the analysis.

For example, in the black-white sample of 200 only three items were identified as biased, while in the sample of 1000, 34 items were identified as biased.

Comparisons Across Sample Sizes and Techniques

Stability

Correlations in black-white samples. Table 12 contains Spearman rank correlations of each type of DIF index with itself for black-white comparisons across different sample sizes. In general, correlations were highest between MH indices in samples of size 750 and 1000 (r=.74) and size 500 and 1000 (r=.72). Correlations between Rasch indices in samples of 500 and 1000 were next highest (r=.50). In general, stability of MH and Rasch indices in the black-white samples is moderate in the largest black-white samples.

Table 12 about here

Correlations in male-female samples. Table 13 contains similar correlations in male-female samples. In this table correlations were highest for the Rasch index in samples of size 750 and 1000 (r=.68), followed by correlations for Delta Plot indices in samples of 750 and 1000 (r=.60). In general, stability of the Rasch and Delta Plot indices is moderate in the male-female samples.

Table 13 about here

Since we also are interested in whether the same items were identified as biased or unbiased in different sample sizes, we examined proportions of items identified by each DIF technique in the samples of 1000 examinees versus all other sample sizes. The results are expressed as a proportion of agreement and are summarized in Tables 14 (white-black) and 15 (male-female). According to Tables 14 and 15, proportions of agreement for the Delta Plot and Rasch techniques are stable across sample sizes. However, the proportions in black-white samples from the MH technique show large variability, ranging from a low of 0.22 for the sample size of 200 versus 1000 to a high of 0.88 for samples of size of 750 and 1000. Proportions for male-female samples from the MH technique are stable across different sample sizes.

Tables 14 and 15 about here

Agreement

Correlations across DIF techniques. Tables 16 and 17 show correlations between DIF indices from each pair of techniques, within each sample size in black-white and male-female comparison groups. Correlations between Rasch and

Delta Plot indices 0.89 (N=750) and 0.90 (N=500 and 1000) and are larger than all other correlations in black-white comparisons. Correlations between Rasch and Delta Plot indices are .87 (N=500), .88 (N=1000), and .90 (N=750). This result is of particular interest for the MTCS since its items are calibrated within the Rasch model but checked for DIF using the Delta Plot method.

Tables 16 and 17 about here

The correlation between the three parameter DIF index -- the criterion for this study-- and the Rasch DIF index in black-white comparisons for N=1000 is the highest of all correlations with the criterion (r=0.54). The correlation between the three parameter and Rasch DIF index for N=1000 in male-female comparisons is also the highest (r=.51).

Proportions of agreement. Correlations are only a crude measure of how well different techniques agree with three-parameter DIF results. We are also interested in the accuracy with which the DIF techniques identify biased and unbiased items, using items identified by the three parameter model as the criterion. We calculated the proportion of agreement between items identified by the three-parameter DIF index and items identified by each of the other three methods. The results are reported in Tables 18 and 19. Both proportions of items identified as "biased" and "unbiased" (i.e., total hits) and proportions identified as "biased" (true positives) are reported. In terms of proportion of total hits, agreement with the three-parameter DIF index is highest for the Delta Plot and Rasch techniques in both black-white and male-female samples. For the MH technique, the proportion of total hits is high in the black-white comparison group (N=200). However, for the black-white samples of 500 or larger, agreement ranges from 0.22 to 0.26. Low MH hit rates were not observed in the male-female samples. We do not discuss proportions of true positives because only three items were identified as potentially racially biased by the three parameter technique.

Tables 18 and 19 about here

Conclusion

The graphical results indicate that no MTCS items are functioning differentially in either black-white or male-female comparisons. Plots of item difficulty estimates for black-white and male-female comparisons show nearly perfect linear relationships in both groups. The patterns of relationships in both race and sex plots are quite similar to the relationships in plots of item difficulty estimates (from Rasch and three parameter calibrations) in the two independent random samples. Patterns are also similar in samples of different sizes.

Differences between ICCs from the three parameter DIF technique identified only three items in the race comparison samples and two items in the sex comparison samples that appear to be functioning differentially. Since confidence intervals around absolute mean differences were used to identify these items, there is a small probability of erroneously detecting items as biased. Because of the potential for false positive errors, and because it can be instructive to identify the features of items that may cause them to function differently in different groups, these items are resubmitted for further judgmental review but are not excluded from test scores.

Stability in black-white samples, as indicated by rank order correlations of the same DIF indices in samples of different sizes, is low tc moderate for the Delta Plot and Rasch DIF techniques. MH stability is high in comparisons of large samples. Stability in male-female samples, as indicated by rank order correlations of the same DIF indices, is moderate for the Rasch, Delta Plot, and MH in large samples.

Proportions of agreement for the Delta Plot and Rasch techniques are stable across sample sizes in white-black and male-female samples. However, agreement proportions from the MH technique in black-white samples show large variablity (i.e , are not stable). Proportion of agreement from the MH technique for male-female samples are stable in different sample sizes.

Agreement, as indicated by rank order correlations across DIF techniques, is very high between Rasch and Delta Plot DIF indices for all sample sizes in both black-white and male-female comparisons. This result is of particular interest since MTCS items are calibrated using the Rasch model but checked for DIF using the Delta Plot method. The Rasch index agrees moderately with the three-parameter DIF index; agreement of other techniques with the three parameter DIF index is low.

In terms of agreement regarding biased and unbiased items, agreement with the three-parameter DIF index is highest for the Delta Plot and Rasch techniques in both black-white and male-female samples. For the MH technique, the proportion of total hits is high in the black-white comparison group of sample size 200. However, for black-white samples of 500 or larger, agreement is low. These findings regarding stability and agreement in real test data partly support previously published research. Harris and

Kolen (1986) Harris and Hoover (1986), and Skaggs and Lissitz (1988) have reported that different DIF techniques do not agree very well with each other and are only moderately stable across different sample sizes. However, these studies have relied mostly on correlations between DIF indices to indicate agreement (cf. Skaggs & Lissitz, 1988); in this study we have reported both correlations and proportions of agreement.

A pattern in chi-square statistics is evident in results from the MH analyses of black-white samples in these data. As the number of response patterns was decreased in white-black samples, chi-square statistics became smaller and identified fewer items as biased. This pattern suggests a dependence between the size of chi-square statistics and sample sizes used in the MH analyses.

We are aware that with the large examinee samples available in statewide testing, chi-square significance tests for item by group interactions using traditional alpha levels may be sensitive to item functioning differences which have no practical importance. In the data used for this study, few non-significant item by group interactions were found in large samples by the MH technique. Proper adjustment of significance levels, so that only practically significant degrees of bias are flagged, requires considerations of effect size and power. However, no attempt was made for adjusting the significance levels in this study.

It should to be mentioned that in previously published research item parameters were estimated using the LOGIST program, while in the present study PC-BILOG was used for three parameter item estimates. BILOG implements marginal maximum likelihood estimation procedures which produce more stable estimates across subgroups than other estimation procedures (see Baghi, 1988).

References

Angoff, W.H. (1972). A technique for the investigation of cultural differences. Paper presented at the annual meeting of the American Psychological Association, Honolulu.

Angoff, W.H., & Ford, S.F. (1973). Item-race interaction on a test of scholastic aptitude. Journal of Educational Measurement, 10, 95-105.

Angoff, W.H. (1982). Use of difficulty and discrimination indices for detecting item bias. In R.A. Berk (Ed.), Handbook of methods for detecting test bias. Baltimore, MD: Johns Hopkins University Press.

Baghi, H. (1988). The stability of item parameter estimates and item characteristic curves across time using different item response models and different estimation procedures. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Burrill, L.E. (1982). Comparative studies of item bias methods. In R.A. Berk (Ed.), Handbook of Methods for Detecting Test Bias. Baltimore: Johns Hopkins University Press.

Cleary, T.A., & Hilton, T.L. (1968). An investigation of item bias. Educational and Psychological Measurement, 28, 61-75.

Cox, D. R. (1970). Analysis of Binary Data. London: Methuen and Co., Ltd.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. NY: Holt, Rinehart and Winston.

Hills, J.R., and Maruhnich, N. (1977). Test item bias analysis: An analysis of methodology and recommendations. Unpublished monograph. Florida State University.

Hills, J.R. (1981). Test item bias detection methods: 1977 to 1980. Unpublished monograph. Florida State University.

Hills, J.R., and King, F.J. (1982). Detecting biased items in minimal competency tests. Paper presented to the Fifth International Symposium on Educational Testing, Stirling, Scotland.

Hambleton, R.K., and Murray, L. (1983). Some goodness of fit investigation for item response models. In R.K. Hambleton (ed.), Applications of item response theor Vancouver: Educational Research Institute of British Columbia.

Harris, D.J., and Hoover, H.D. (1986). The stability of selected item bias indices. Paper presented at the National Council on Measurement in Education, San Francisco.

Harris, D.J., and Kolen, M.J. (1986). A comparison of the stability of item bias indices using the bootstrap. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Holland, P.W., & Thayer, D.T. (1986). Differential item performance and the Mantel-Haenszel procedure. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Holland, P. W., & Thayer, D.T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H.I. Braun (Eds.), Test validity. Hillsdale, NJ: Lawrence Erlbaum Associates.

Ironson, G. H. (1982). Use of chi-square and latent trait approaches for detecting item bias. In R. A. Berk (Ed.), Handbook of Methods for Detecting Test Bias. Baltimore: Johns Hopkins University Press.

Lord, F.M. (1977). A study of item bias using item characteristic curve theory. In Y. H. Poortinga (Ed.), Bias problems in cross-cultural psychology. Amsterdam: Swets and Zeitlinger.

Mislevy, R.J., & Bock, R.D. (1986). PC-Bilog. [Computer program.] Moorseville, IN: Scientific Software.

Osterlind, S.L. (1987). Test item bias: Quantitative applications in social sciences. Beverly Hills, CA: Sage Publications.

Phillips S.E., and Mehrens, W.A. (1988). Comparison of methods for detecting differential item performance due to instructional test misalignment. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Raju, N.S., Bode, R.K., & Larsen, V. S. (1989). Am empirical assessment of the Mantel-Haenszel statistic for studying differential item performance. Applied Measurement on Education, 2 (1), 1-13.

Reckase, M.D. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. Journal of Educational Statistics, 4, 207-230.

Rudner, L. M. (1977). An evaluation of select approaches for biased item modification. Unpublished doccoral dissertation, Catholic University of America, Washington, D.C.

Rudner, L.M., Getson, P.R., & Knight, D.L. (1980). Biased item detection techniques. Journal of Educational Statistics, 5, 213- 233.

Scheuneman, J.D. (1979). A method of assessing bias in test items. Journal of Educational Measurement, 16, 143-152.

Shepard, L. A., Camilli G., and Williams, D. (1985). Validity of approximation techniques for detecting item bias. Journal of Educational Measurement, 22, 77-105.

Skaggs, G., & Lissitz, R. (1988). The effect of estimation methods on the stability of selected item bias indices. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.

Subkoviak, M.J., Mack, J.S., Ironson, G.H. and Craig, R.D. (1984). Empirical comparison of selected item detection procedures with bias manipulation. Journal of Educational Measurement, 21, 49-58.

Wright, B.D., Mead, R., & Bell, S. (1979). BICAL: Calibrating items with the Rasch model (Report No. RM23b). Chicago: Statistical Laboratory, Department of Education, University of Chicago.

stevef.doc

Table 1

Descriptive Statistics For All Student Subsamples

Samples	N	Mean	Median	SD	P	rpbis	KR-20
Random 1	200	32.87	34.00	7.49	0.73	0.39	0.84
Random 2	200	33.49	35.00	8.34	0.74	0.42	0.85
Random 3	500	32.97	34.00	8.21	0.73	0.42	0.85
Random 4	500	33.49	36.00	8.58	0.74	0.40	0.86
Random 5	750	33.49	36.00	8.35	0.74	0.45	0.85
Random 6	750	33.59	36.00	8.47	0.75	0.39	0.85
Random 7	1000	33.09	35.00	8.53	0.74	0.42	0.86
Random 8	1000	33.28	35.00	8.48	0.74	0.45	0.86
White 1	200	35.84	38.00	7.43	0.80	0.39	0.82
White 2	500	35.23	37.00	7.81	0.78	0.42	0.83
White 3	750	35.05	37.00	7.18	0.78	0.46	0.82
White 4	1000	35.12	37.00	7.49	0.78	0.41	0.83
Black 1	200	30.91	32.00	8.61	0.69	0.44	0.87
Black 2	500	30.79	31.00	8.04	0.68	0.45	0.85
Black 3	750	31.07	32.00	8.45	0.69	0.42	0.86
Black 4	1000	31.04	32.00	8.18	0.69	0.39	0.86
Male 1	200	33.51	36.00	9.06	0.74	0.43	0.86
Male 2	500	33.26	35.00	8.82	0.74	0.41	0.86
Male 3	750	34.14	36.00	8.23	0.76	0.44	0.84
Male 4	1000	33.75	36.00	8.54	0.75	0.44	0.85
Female 1	200	33.04	33.00	7.63	0.73	0.40	0.84
Female 2	500	33.97	35.00	7.86	0.75	0.42	0.84
Female 3	750	33.69	35.00	7.85	0.75	0.45	0.84
Female 4	1000	33.71	35.00	7.84	0.75	0.41	0.84

Table 2

Item Parameter Estimates For Selected Student Subsamples (N=1000)

	Random1			Random2			White			Black			Male			Female		
Item #	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c	a	b	c
1	0.498	-3.393	0.267	0.741	-2.421	0.290	0.735	-2.612	0.268	0.631	-2.363	0.253	0.815	-2.463	0.273	1.043	-1.771	0.105
2	0.847	0.527	0.253	0.606	0.193	0.213	0.757	0.491	0.287	0.635	0.598	0.243	0.846	0.311	0.311	0.579	0.169	0.096
3	0.958	-1.125	0.277	0.894	-1.104	0.234	0.788	-1.604	0.315	0.735	-1.176	0.241	0.861	-1.481	0.222	1.008	-1.444	0.089
4	0.811	0.292	0.242	0.937	0.477	0.278	0.994	0.181	0.275	0.960	0.660	0.307	0.801	0.361	0.306	0.652	-0.037	0.133
5	0.468	0.404	0.220	0.491	0.503	0.201	0.536	0.690	0.287	0.472	0.680	0.211	0.544	0.439	0.247	0.462	0.045	0.105
6	1.444	-0.326	0.337	1.142	-0.519	0.293	1.207	-0.792	0.314	0.969	-0.048	0.359	0.973	-0.620	0.342	1.081	-0.702	0.143
7	0.818	-2.008	0.317	0.973	-1.863	0.350	0.935	-2.265	0.283	0.832	-1.754	0.291	0.866	-2.219	0.282	0.923	-1.828	0.105
8	0.730	0.310	0.153	0.668	0.099	0.112	0.804	0.203	0.245	0.855	0.540	0.179	0.720	-0.126	0.112	0.746	0.101	0.096
9	1.469	0.035	0.164	1.617	0.144	0.230	1.417	-0.275	0.155	1.398	0.310	0.216	1.529	-0.267	0.200	1.386	0.068	0.133
10	1.708	-0.153	0.441	1.483	-0.276	0.418	1.157	-0.687	0.304	1.235	-0.055	0.452	1.314	-0.580	0.381	1.027	-0.261	0.108
11	1.310	0.305	0.175	0.975	0.439	0.168	1.096	0.227	0.173	1.123	0.594	0.189	1.266	0.404	0.220	0.986	0.189	0.095
12	1.376	-0.156	0.442	1.322	-0.171	0.474	1.026	-0.452	0.362	0.789	-0.235	0.445	0.882	-0.759	0.264	0.819	-0.920	0.109
13	0.776	-0.393	0.223	0.724	-0.525	0.197	0.795	-0.628	0.201	1.007	-0.057	0.183	0.677	-0.594	0.187	0.691	-0.555	0.081
14	1.416	0.551	0.256	1.150	0.551	0.261	0.974	0.453	0.343	1.191	0.891	0.231	1.269	0.423	0.288	0.752	0.296	0.153
15	0.933	0.488	0.244	0.792	0.428	0.230	1.026	0.359	0.279	0.795	0.677	0.240	0.952	0.173	0.256	0.759	0.335	0.125
16	0.803	-1.061	0.279	0.883	-1.146	0.275	0.822	-1.713	0.234	1.129	-0.266	0.474	0.794	-1.381	0.250	0.928	-1.135	0.104
17	1.009	-0.608	0.292	1.190	-0.464	0.389	0.837	-1.106	0.292	0.763	-0.458	0.257	1.017	-0.753	0.345	1.066	-0.882	0.094
18	0.804	-0.534	0.188	0.796	-0.668	0.185	0.780	-0.930	0.218	0.658	-0.198	0.183	0.841	-0.577	0.222	0.917	-0.252	0.103
19	1.209	0.588	0.148	0.907	0.415	0.151	0.750	0.234	0.161	0.969	1.047	0.220	0.890	0.261	0.127	0.954	0.470	0.138
20	1.230	0.420	0.280	1.200	0.265	0.208	0.890	-0.093	0.213	1.124	0.683	0.235	0.920	-0.074	0.158	1.199	0.291	0.198
21	0.606	-1.458	0.201	0.552	-1.397	0.236	0.408	-2.043	0.267	0.432	-1.297	0.258	0.629	-1.331	0.233	0.638	-1.389	0.107
22	1.173	-1.196	0.264	1.253	-1.283	0.181	1.214	-1.513	0.273	1.282	-0.819	0.273	1.128	-1.229	0.185	1.505	-1.140	0.099
23	0.589	-1.889	0.206	0.596	-2.036	0.242	0.465	-2.518	0.244	0.503	-2.115	0.231	0.564	-1.989	0.187	0.708	-1.909	0.106
24	0.984	-2.406	0.202	0.919	-2.756	0.228	0.640	-3.674	0.243	0.927	-2.367	0.223	0.930	-2.305	0.213	1.433	-2.020	0.089
25	0.830	-1.315	0.240	0.682	-1.497	0.199	0.683	-2.077	0.199	0.556	-1.157	0.221	0.858	-1.614	0.217	0.961	-1.252	0.096
26	0.874	-0.781	0.152	0.957	-0.701	0.168	0.786	-1.367	0.179	0.993	-0.357	0.229	0.837	-1.193	0.153	0.977	-0.797	0.093
27	0.780	-1.942	0.187	0.654	-2.178	0.220	0.846	-2.390	0.204	0.703	-1.863	0.211	0.666	-2.193	0.212	1.093	-1.808	0.094
28	1.356	-1.821	0.178	1.378	-1.811	0.196	1.018	-2.413	0.217	0.814	-2.060	0.205	1.129	-2.019	0.169	1.242	-1.775	0.086
29	0.999	-1.101	0.221	1.055	-1.101	0.196	0.966	-1.546	0.218	0.966	-0.934	0.213	0.877	-1.301	0.240	1.105	-1.153	0.084
30	0.744	-1.227	0.172	0.791	-1.084	0.209	0.704	-1.616	0.210	0.618	-0.906	0.206	0.890	-1.115	0.162	0.919	-1.067	0.091
31	0.955	-1.998	0.201	0.603	-2.402	0.235	0.681	-2.605	0.216	1.145	-1.650	0.208	0.849	-2.124	0.178	0.879	-1.796	0.095
32	0.754	-1.015	0.176	0.667	-0.934	0.238	0.924	-1.155	0.215	0.610	-0.826	0.249	0.764	-1.164	0.214	0.812	-1.055	0.095
33	0.993	-1.642	0.185	0.831	-1.811	0.178	0.948	-2.147	0.200	0.969	-1.555	0.236	1.081	-1.706	0.171	1.277	-1.448	0.083
34	0.555	-0.017	0.199	0.497	0.045	0.257	0.610	-0.452	0.160	0.628	0.191	0.213	0.753	-0.052	0.250	0.578	-0.184	0.095
35	0.783	-0.762	0.188	0.874	-0.529	0.198	0.972	-0.731	0.266	1.060	-0.187	0.251	0.905	-0.522	0.227	0.902	-0.934	0.096
36	1.410	-0.868	0.203	1.736	-0.631	0.294	1.594	-0.956	0.266	1.679	-0.490	0.334	1.371	-0.998	0.160	1.682	-0.705	0.094
37	0.988	-0.318	0.173	1.292	0.005	0.267	1.471	-0.127	0.325	1.307	-0.008	0.234	1.160	-0.401	0.216	1.266	-0.324	0.129
38	1.020	-0.724	0.243	1.109	-0.629	0.284	1.123	-0.967	0.290	1.046	-0.504	0.273	1.258	-0.790	0.254	1.154	-0.783	0.098
39	0.886	-0.716	0.241	1.093	-0.530	0.251	0.816	-0.993	0.203	0.918	-0.257	0.270	0.868	-0.758	0.192	0.982	-0.838	0.093
40	0.905	-0.335	0.155	1.145	-0.144	0.233	1.331	-0.366	0.254	1.014	0.017	0.171	1.074	-0.282	0.180	1.021	-0.327	0.124
41	1.580	-0.049	0.201	1.348	-0.102	0.212	1.453	-0.315	0.215	1.533	0.157	0.250	1.708	-0.087	0.207	1.173	-0.333	0.082
42	1.375	-0.456	0.233	1.398	-0.579	0.220	1.499	-0.735	0.281	1.515	-0.298	0.292	1.518	-0.736	0.208	1.472	-0.648	0.091
43	1.377	-0.189	0.095	1.148	-0.282	0.088	1.352	-0.260	0.167	1.318	0.071	0.167	1.354	-0.141	0.158	1.116	-0.237	0.079
44	0.619	-0.219	0.262	0.585	-0.311	0.199	0.584	-0.475	0.241	0.457	-0.067	0.202	0.636	-0.233	0.243	0.557	-0.747	0.100
45	0.665	-1.315	0.203	0.593	-1.473	0.213	0.665	-1.612	0.212	0.601	-1.156	0.244	0.595	-1.497	0.217	0.810	-1.330	0.092

Note. a = discrimination; b = difficulty level; c = guessing.

Table 3

Means, Standard Deviations, and Confidence Intervals for Differences Between ICCs for Rasch and Three-Parameter Models: Within Subsamples Random 1 vs. Random 2, White vs. Black, Male vs. Female.

	3 Parameter (N=1000)			Rasch (N=200)			Rasch (N=500)			Rasch (N=750)			Rasch (N=1000)		
	R1-R2	W-B	M-F	R1-R2	W-B	M-F	R1-R2	W-B	M-F	R1-R2	W-B	M-F	R1-R2	W-B	M-F
Mean of Diff.	-0.01591	0.01207	0.25347	0.00653	0.00167	0.00478	0.00145	0.00408	-0.00269	0.00062	0.00224	0.00176	0.00207	0.00277	0.00102
SD of Diff.	0.09141	0.14032	0.17028	0.31209	0.36481	0.29028	0.26562	0.25541	0.26804	0.15001	0.27396	0.23097	0.12217	0.22913	0.23383
C.I. of Diff.	-0.25725	-0.35839	-0.19608	-0.81738	-0.96143	-0.76156	-0.69978	-0.67021	-0.71032	-0.3954	-0.72101	-0.60799	-0.32046	-0.60214	-0.61630
	0.22542	0.38252	0.70302	0.83044	0.96477	0.77112	0.70268	0.67837	0.70495	0.39664	0.7255	0.61152	0.32459	0.60767	0.61834
Mean of Abs. Diff.	0.12583	0.16558	0.28624	0.25049	0.3274	0.2941	0.19661	0.19944	0.21573	0.1144	0.24605	0.18631	0.10161	0.20034	0.19570
SD of Abs. Diff.	0.05557	0.08599	0.14008	0.18255	0.19661	0.12196	0.17902	0.15932	0.15878	0.0959	0.14784	0.14058	0.06641	0.11867	0.12854
C.I. of Abs. Diff.	-0.02087	-0.06142	-0.08358	-0.23143	-0.19164	-0.02787	-0.27599	-0.22117	-0.20345	-0.13878	-0.14423	-0.18483	-0.07372	-0.11296	-0.14364
	0.27254	0.39259	0.65606	0.73241	0.84644	0.61607	0.66922	0.62004	0.63491	0.36758	0.63634	0.55745	0.27694	0.51363	0.53504

Note. R1-R2 = difference between random samples 1 and 2; W-B = difference between white and black samples; M-F = difference between male and female samples.

FIGURE 1 Plot of Item Difficulty Estimates from the Three-Parameter Model in Two Random Samples. (N=1000)

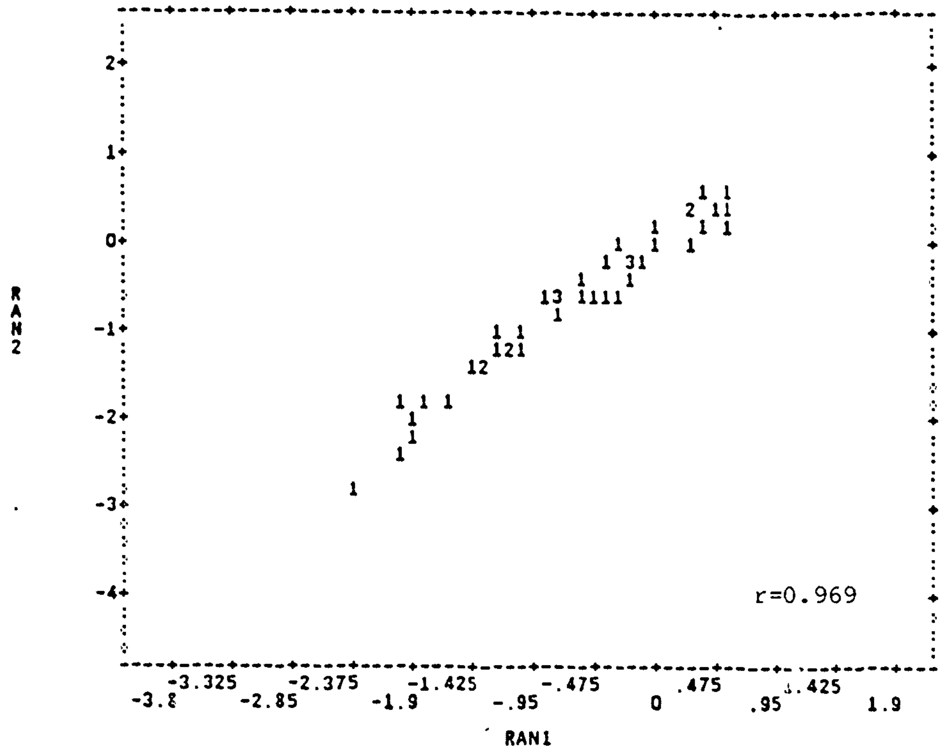

FIGURE 2 Plot of Item Discrimination Estimates from the Three-Parameter Model in Two Random Samples. (N=1000)

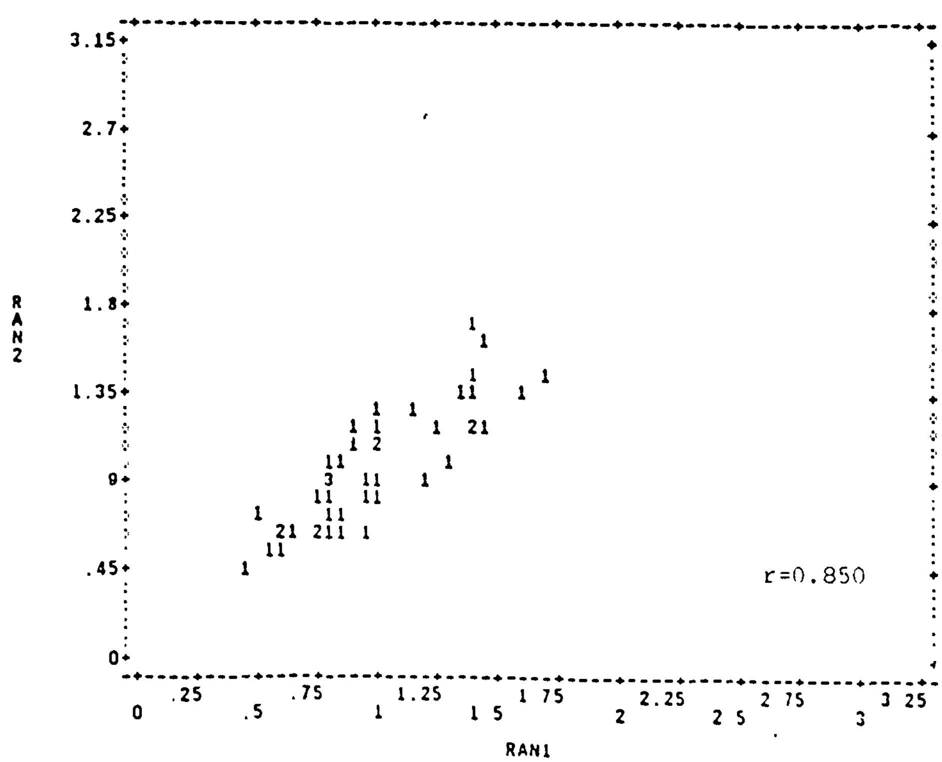

FIGURE 3 Plot of Item Difficulty Estimates from the Three Parameter Model In Black and White Samples. (N=1000)

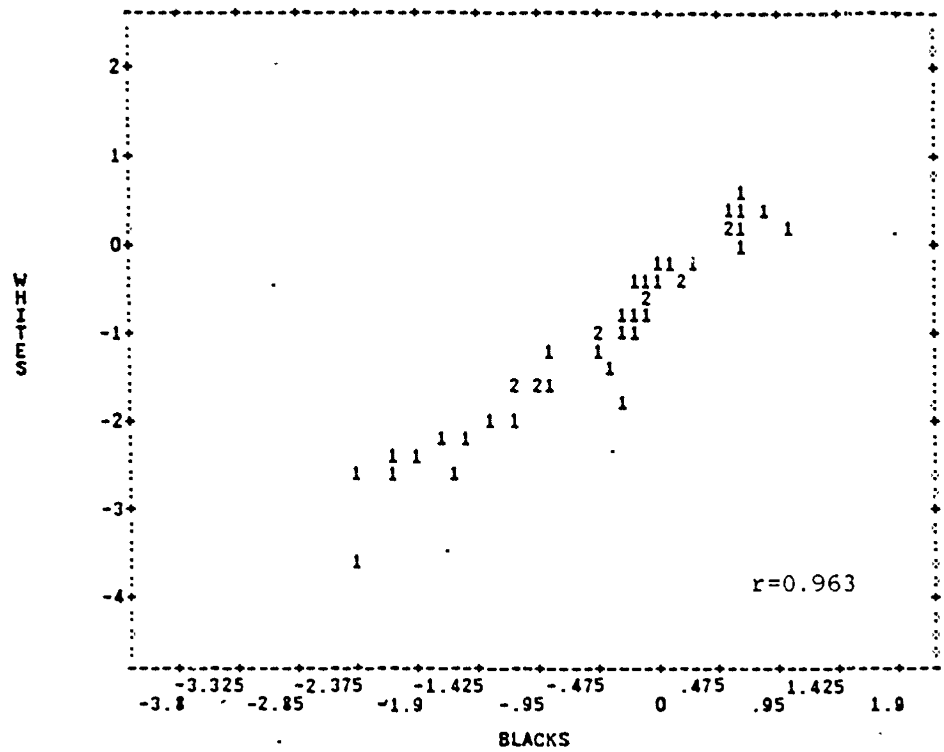

FIGURE 4 Plot of Item Discrimination Estimates from the Three-Parameter Model in Black and White Samples. (N=1000)

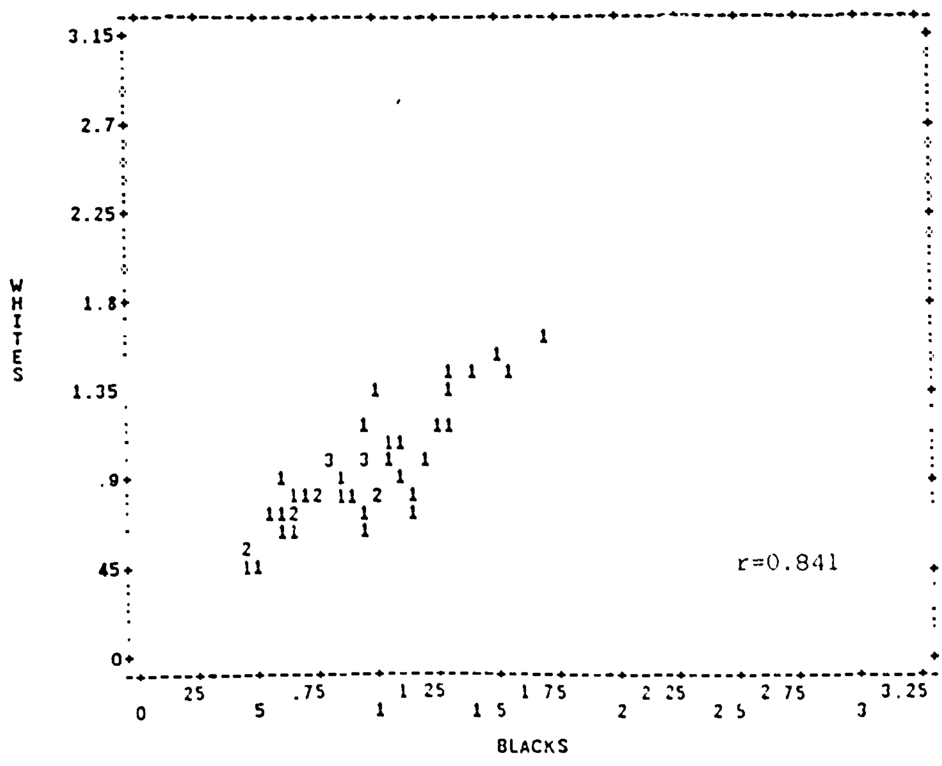

Figure 5 Plot of Item Difficulty Estimates from the Three-Parameter Model in Male and Female Samples. (N=1000)

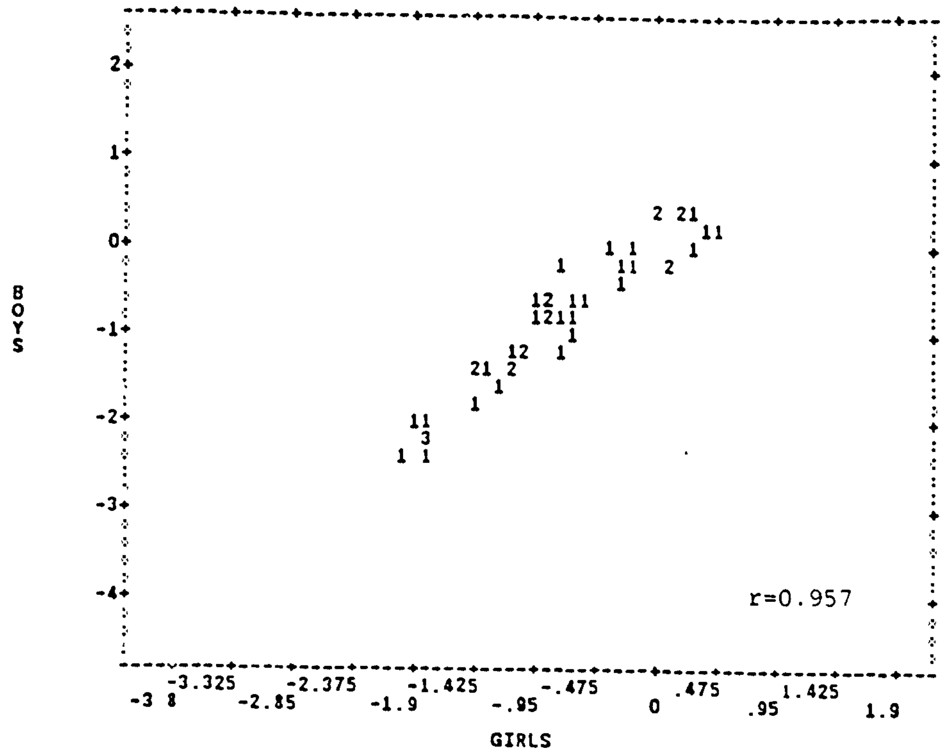

FIGURE 6 Plot of Item Discrimination Estimates from the Three-Parameter Model in Male and Female Samples. (N=1000)

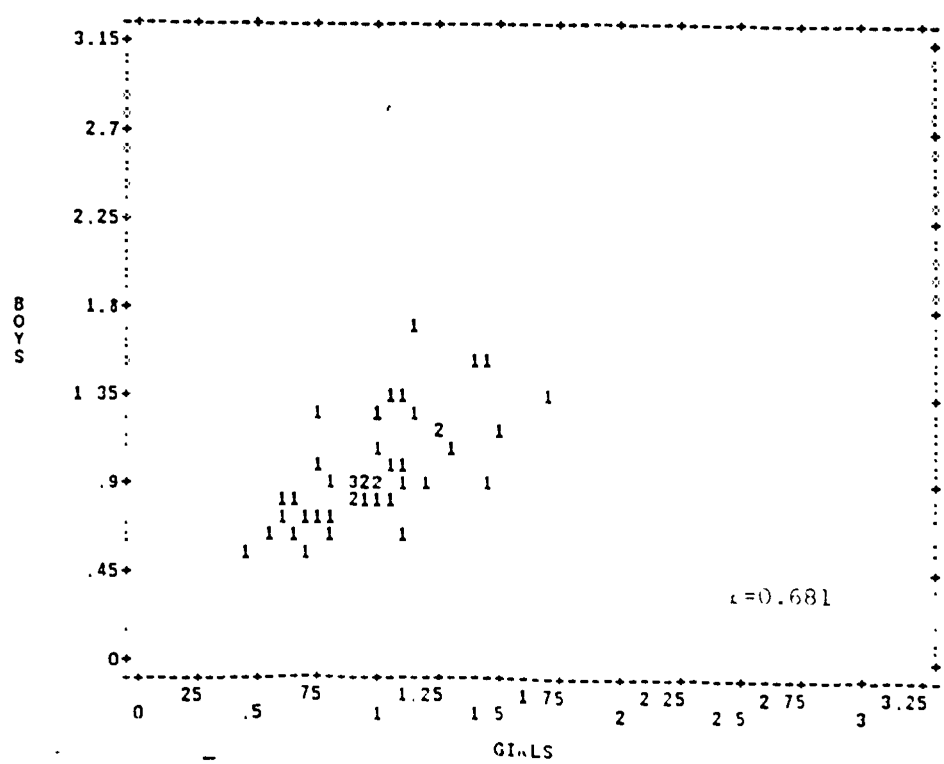

FIGURE 7 Plot of Item Difficulty Estimates from the Rasch Model in Two Random Samples. (N=200)

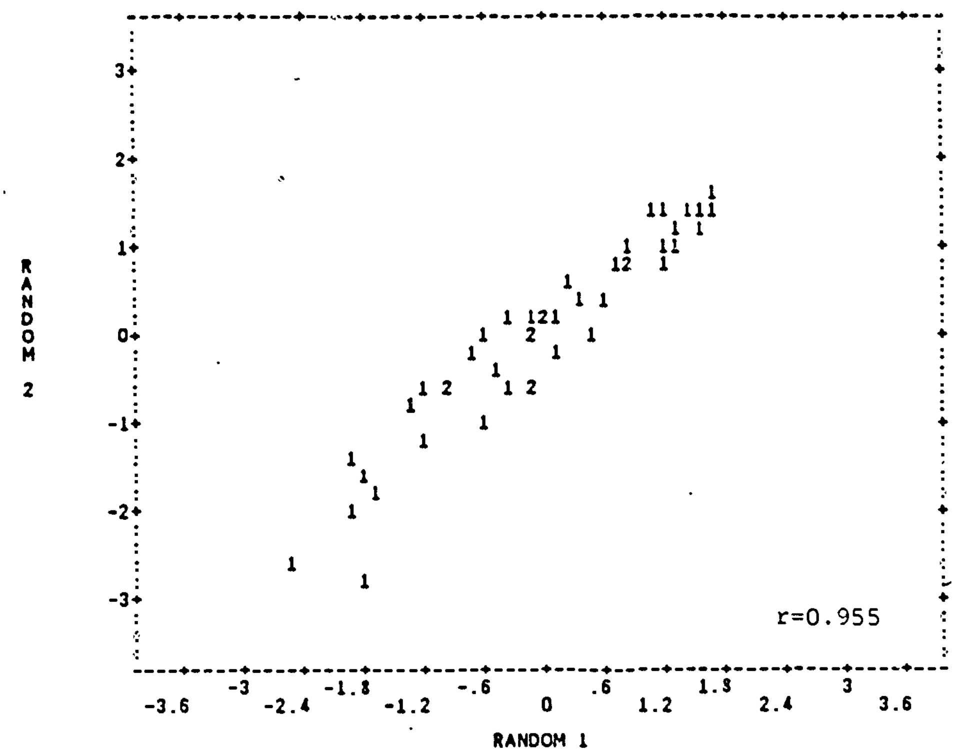

FIGURE 8 Plot of Item Difficulty Estimates from the Rasch Model in Black and White Samples. (N=200)

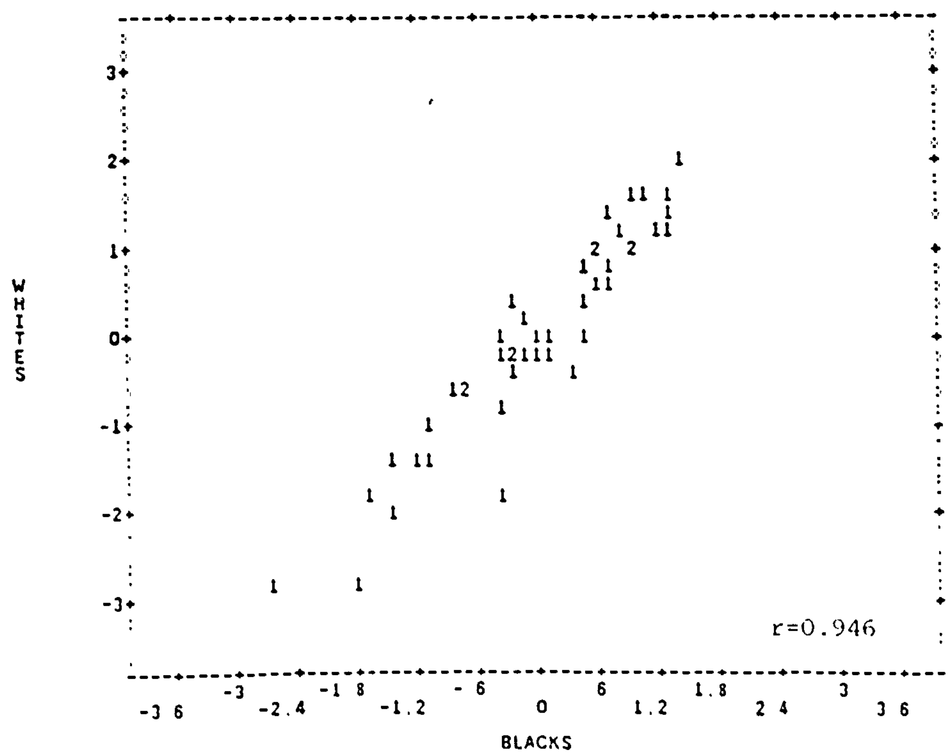

FIGURE 9 Plot of Item Difficulty Estimates from the Rasch Model in Two Random Samples. (N=200)

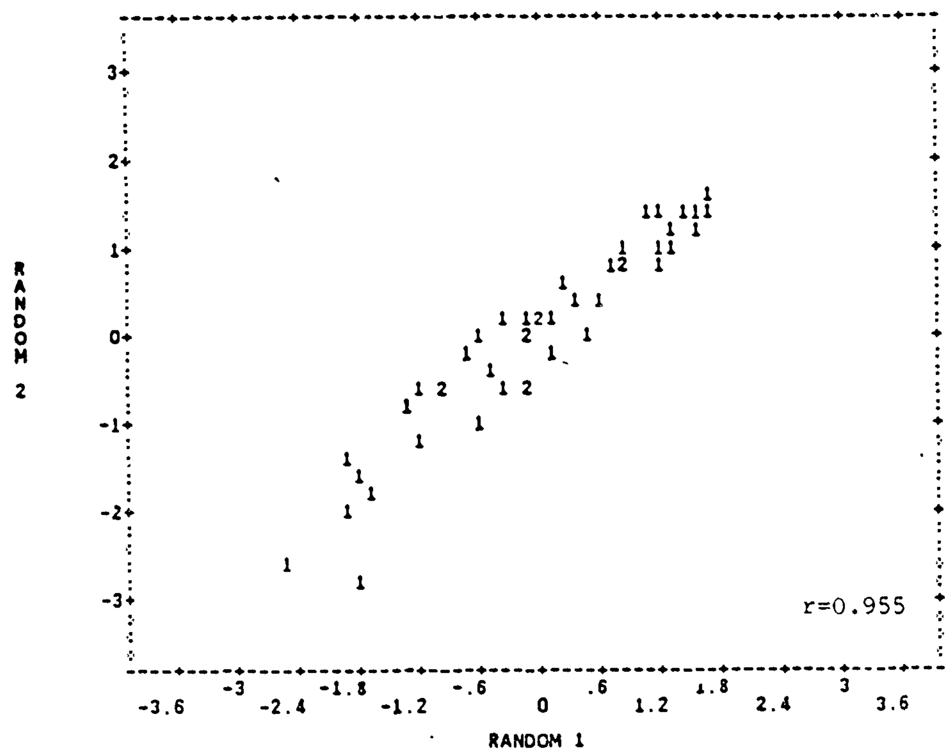

FIGURE 10 Plot of Item Difficulty Estimates from the Rasch Model in Male and Female Samples. (N=200)

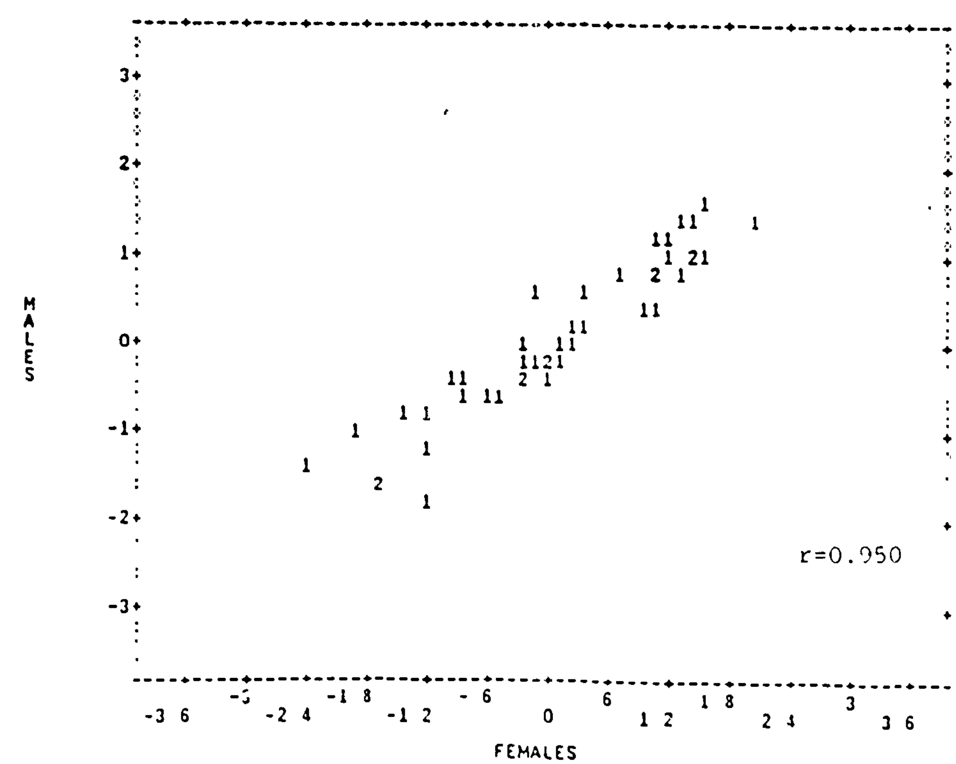

FIGURE 11 Plot of Item Difficulty Estimates from the Rasch Model in Two Random Samples. (N=500)

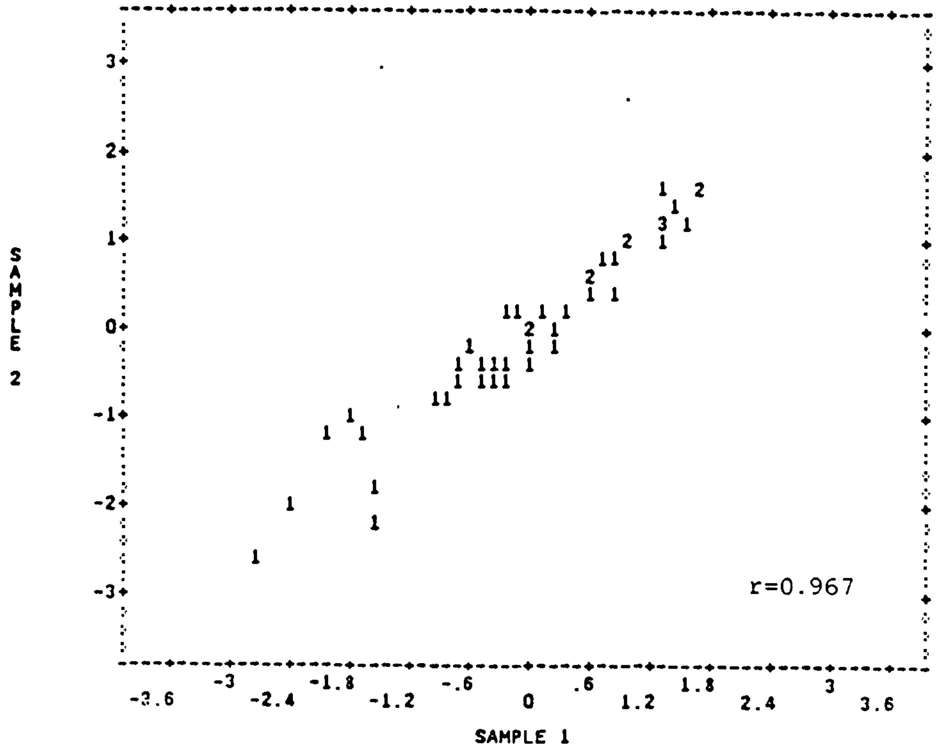

FIGURE 12 Plot of Item Difficulty Estimates from the Rasch Model in Black and White Samples. (N=500)

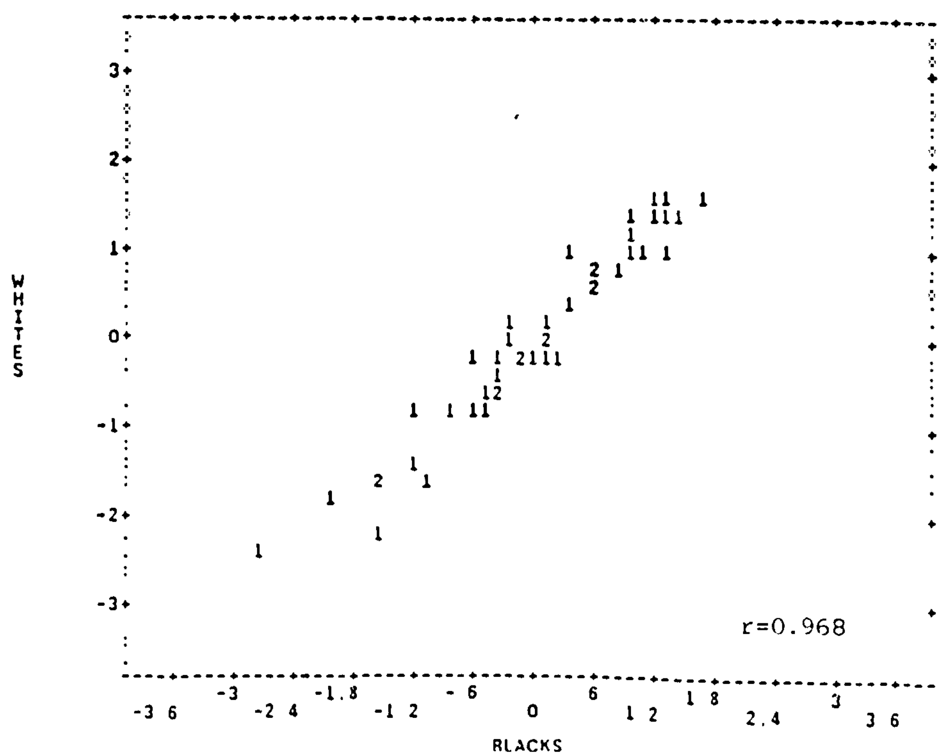

FIGURE 13 Plot of Item Difficulty Estimates from the Rasch Model in Two Random Samples. (N=500)

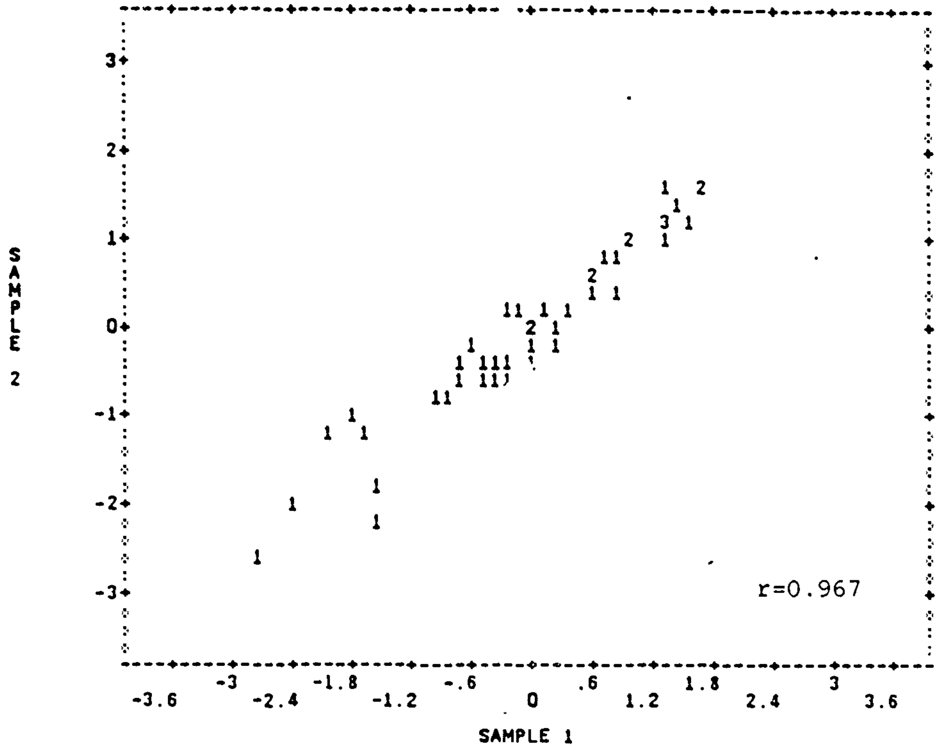

FIGURE 14 Plot of Item Difficulty Estimates from the Rasch Model in Male and Female Samples. (N=500)

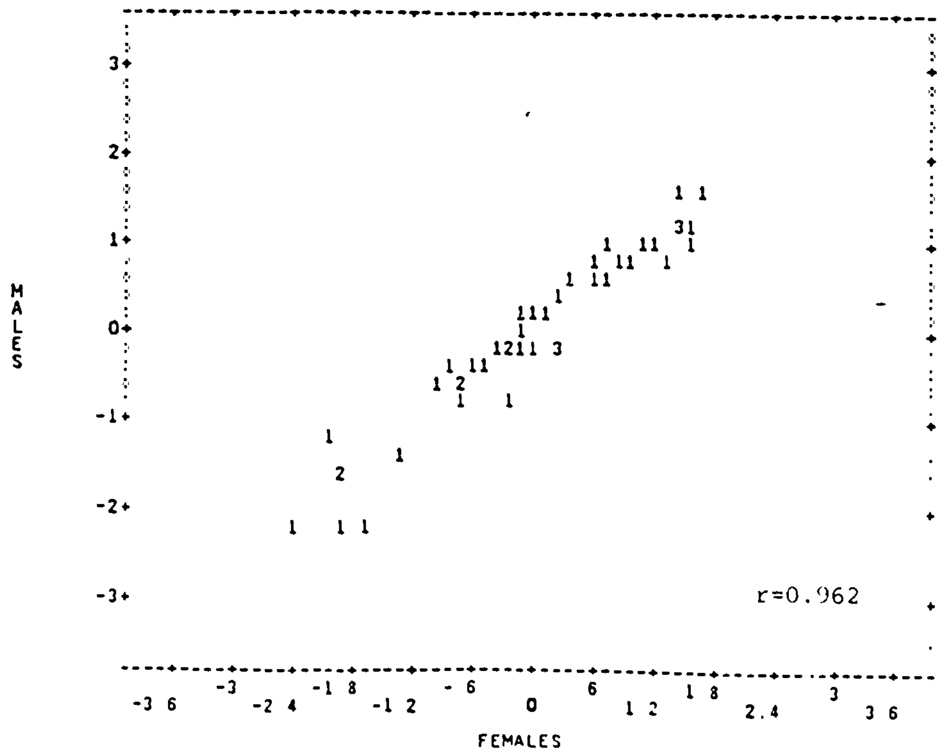

FIGURE 15 Plot of Item Difficulty Estimates from the Rasch Model in Two Random Samples. (N=750)

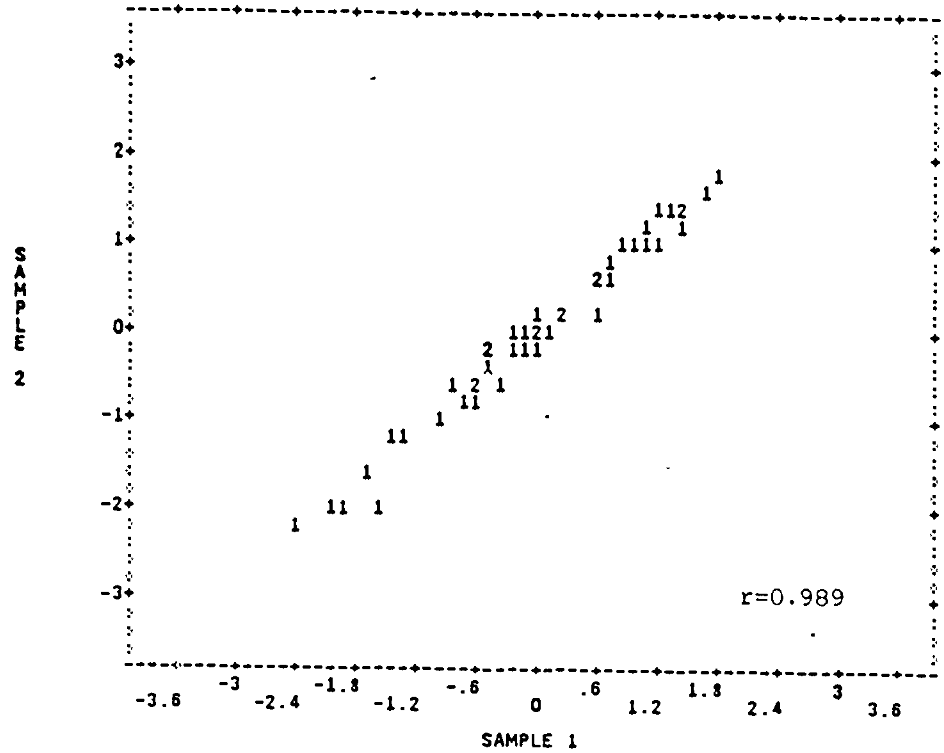

FIGURE 16 Plot of Item Difficulty Estimates from the Rasch Model in Black and White Samples. (N=750)

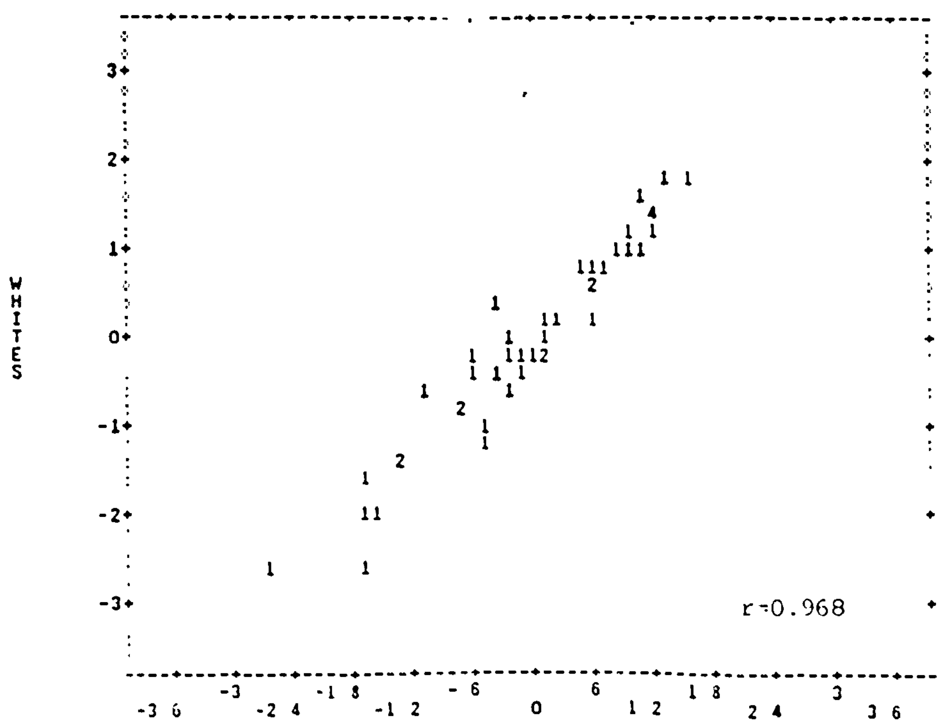

FIGURE 17 Plot of Item Difficulty Estimates from the Rasch Model in Two Random Samples. (N=750)

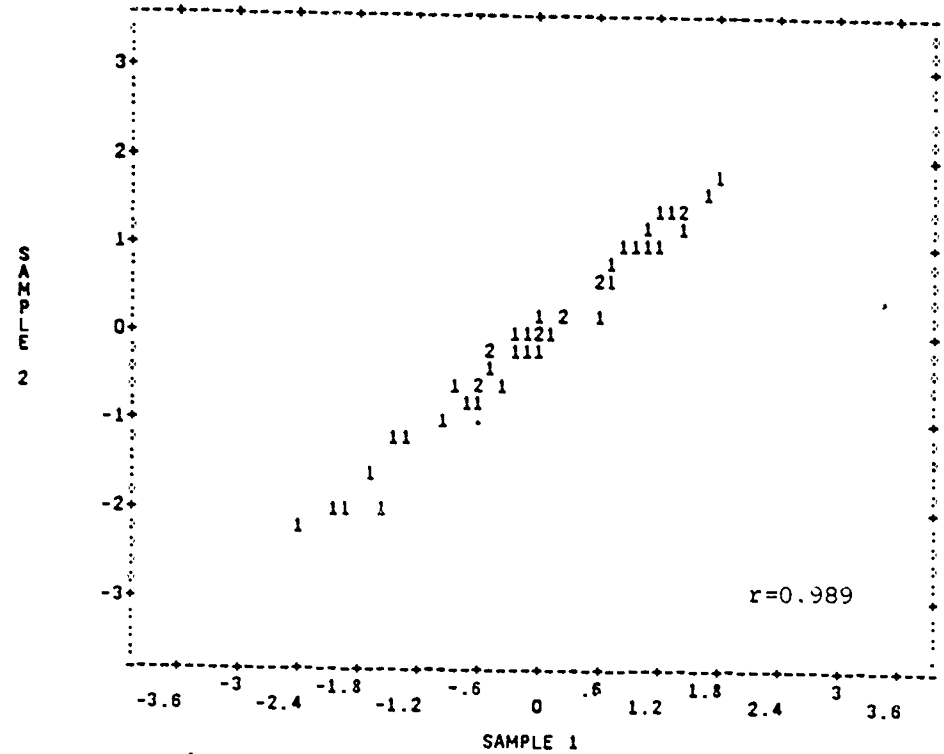

FIGURE 18 Plot of Item Difficulty Estimates from the Rasch Model in Male and Female Samples. (N=750)

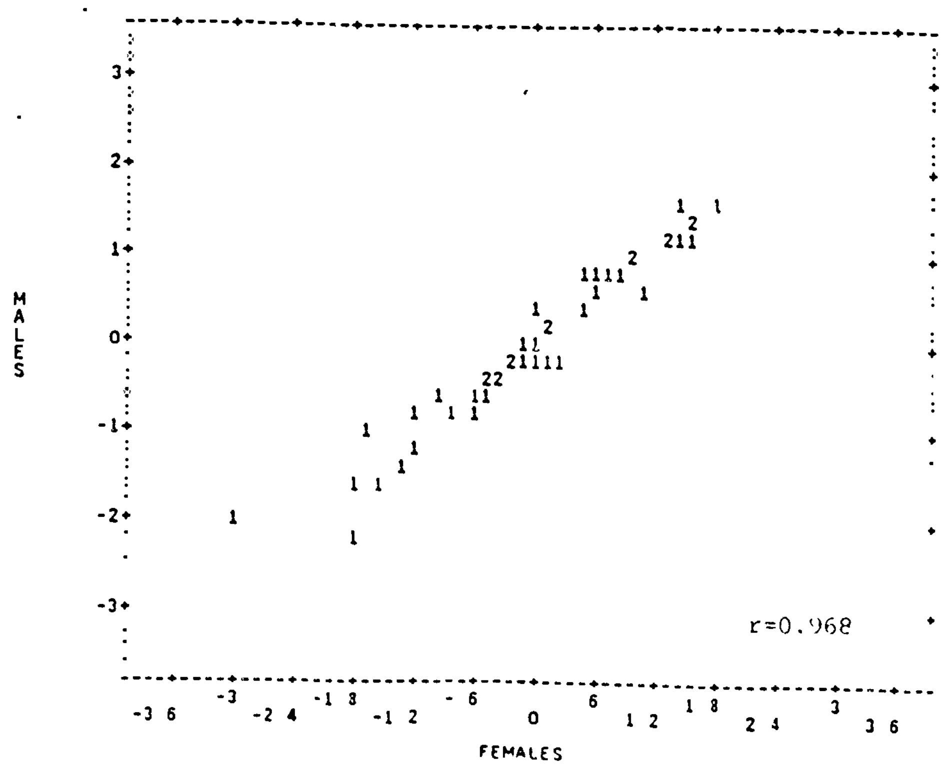

FIGURE 19 Plot of Item Difficulty Estimates from the Rasch Model in Two Random Samples. (N=1000)

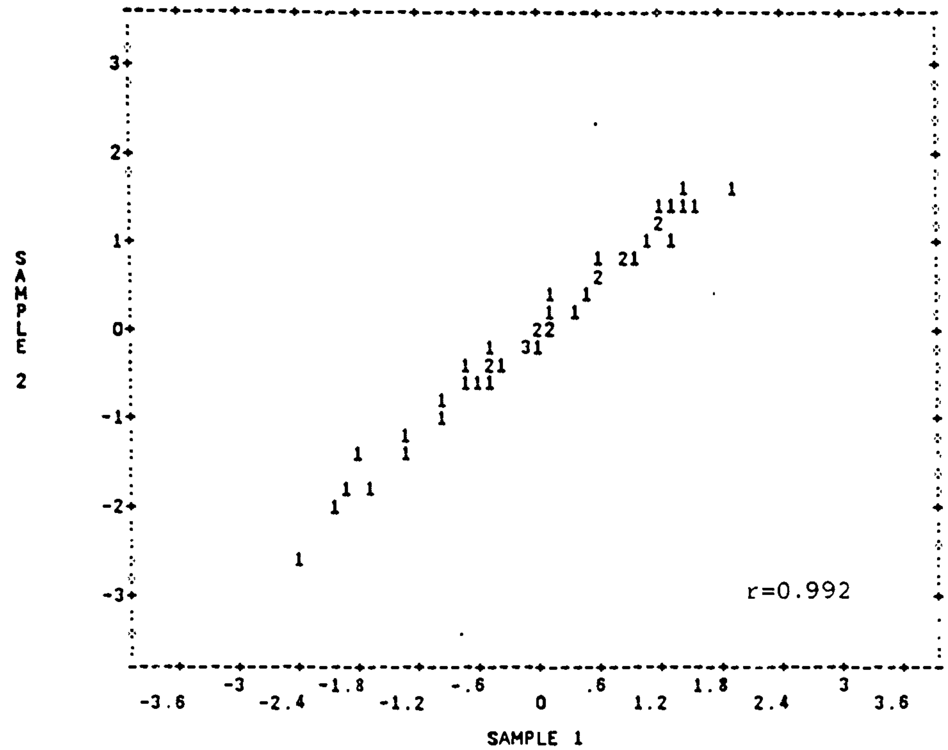

FIGURE 20 Plot of Item Difficulty Estimates from the Rasch Model in Black and White Samples. (N=1000)

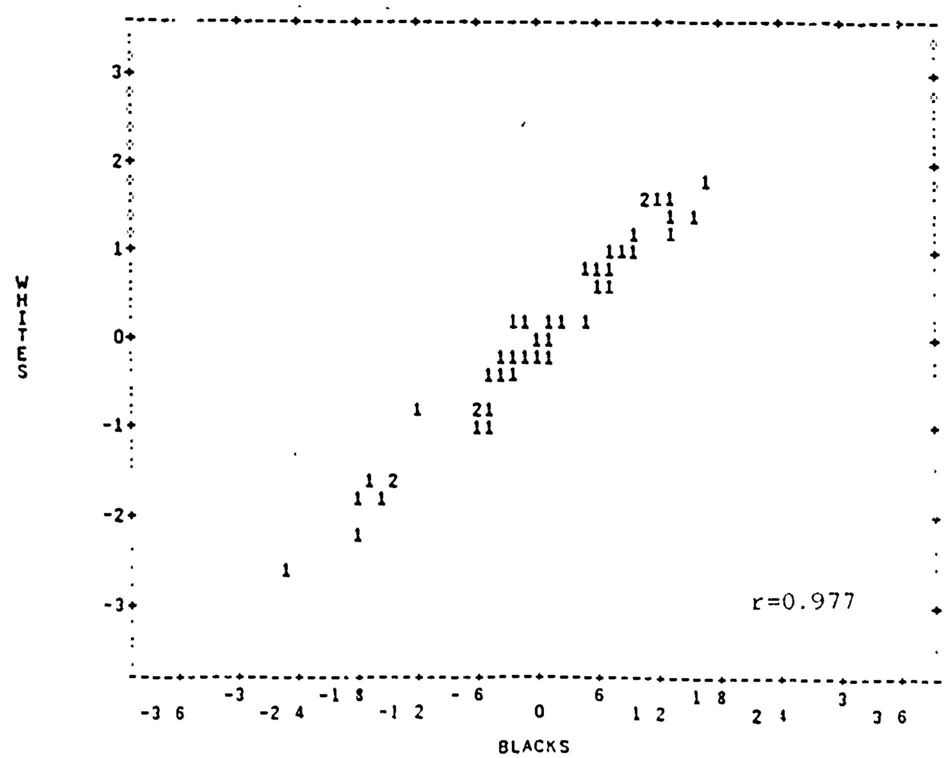

FIGURE 21 Plot of Item Difficulty Estimates from the Rasch Model in Two Random Samples. (N=1000)

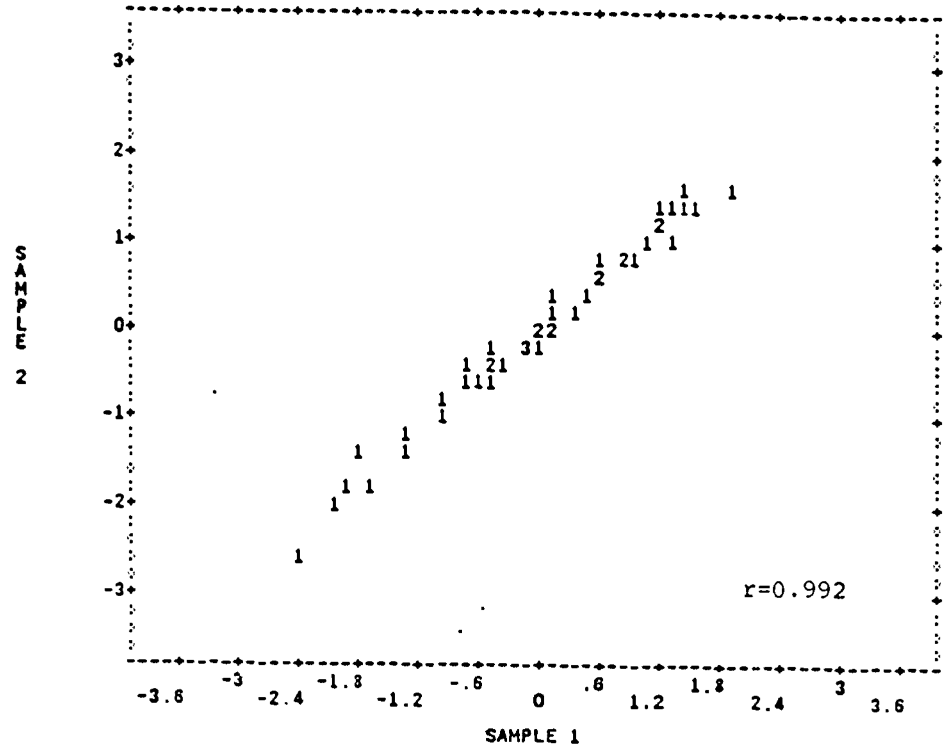

FIGURE 22 Plot of Item Difficulty Estimates from the Rasch Model in Male and Female Samples. (N=1000)

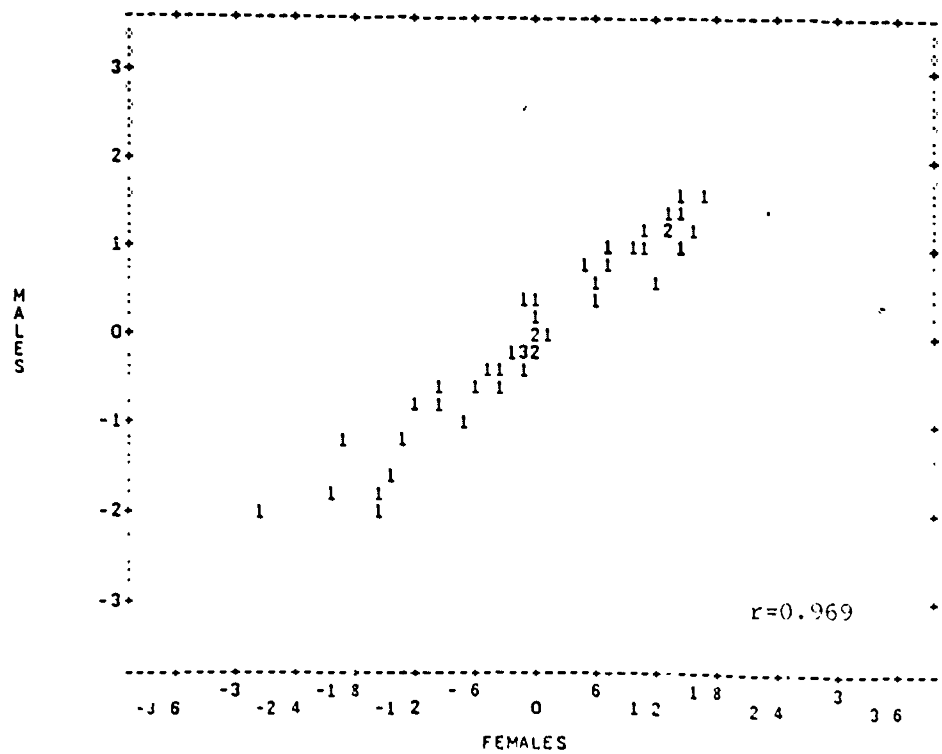

TABLE 4

Item Statistics and DIF Indices for the Delta Plot Method: White VS Black. (N = 200)

NUMBER	P-VALUE REF	P-VALUE FOC	Z-VALUE REF	Z-VALUE FOC	DELTA REF	DELTA FOC	BIAS*
1	.96	.90	1.75	1.28	20.00	18.12	-.93
2	.53	.54	.08	.10	13.32	13.40	.62
3	.86	.87	1.08	1.13	17.32	17.52	.59
4	.56	.56	.15	.15	13.60	13.60	.55
5	.56	.49	.15	-.03	13.60	12.88	.06
6	.78	.68	.77	.47	16.08	14.88	-.35
7	.95	.91	1.65	1.34	19.60	18.36	-.47
8	.60	.47	.25	-.08	14.00	12.68	-.37
9	.64	.50	.36	.00	14.44	13.00	-.47
10	.77	.80	.74	.84	15.96	16.36	.76
11	.52	.45	.05	-.13	13.20	12.48	.07
12	.73	.75	.61	.67	15.44	15.68	.67
13	.72	.64	.58	.36	15.32	14.44	-.10
14	.54	.49	.10	-.03	13.40	12.88	.20
15	.54	.56	.10	.15	13.40	13.60	.70
16	.88	.75	1.17	.67	17.68	15.68	-.95
17	.79	.72	.81	.58	16.24	15.32	-.16
18	.74	.69	.64	.50	15.56	15.00	.11
19	.51	.43	.03	-.18	13.12	12.28	-.01
20	.64	.44	.36	-.15	14.44	12.40	-.88
21	.85	.77	1.04	.74	17:16	15.96	-.38
22	.90	.83	1.28	.95	18.12	16.80	-.49
23	86	.84	1.08	.99	17.32	16.96	.20
24	.94	.96	1.55	1.75	19.20	20.00	.95
25	.89	.83	1.23	.95	17.92	16.80	-.34
26	.84	.72	.99	.58	16.96	15.32	-.68
27	.95	.89	1.65	1.23	19.60	17.92	-.78
28	.94	.95	1.55	1.65	19.20	19.60	.67
29	.83	.82	.95	.92	16.80	16.68	.38
30	.86	.76	1.08	.71	17.32	15.84	-.58
31	.89	.91	1.23	1.34	17.92	18.36	.74
32	.77	.79	.74	.81	15.96	16.24	.68
33	.89	.90	1.23	1.28	17.92	18.12	.57
34	.58	.55	.20	.13	13.80	13.52	.35
35	.75	.70	.67	.52	15.68	15.08	.08
36	.80	.75	.84	.67	16.36	15.[illegible]8	.01
37	.68	.60	.47	.25	14.88	14.00	-.09
38	.81	.69	.88	.50	16.52	15.00	-.58
39	.80	.70	.84	.52	16.36	15.08	-.41
40	.69	.59	.50	.23	15.00	13.92	-.23
41	.66	.56	.41	.15	14.64	13.60	-.20
42	.75	.71	.67	.55	15.68	15.20	.16
43	.61	.55	.28	.13	14.12	13.52	.12
44	.69	.62	.50	.30	15.00	14.20	-.04
45	.82	.80	.92	.84	16.68	16.36	.25

FIGURE 23 Plot of Delta Values for White and Black Samples. (N=200)

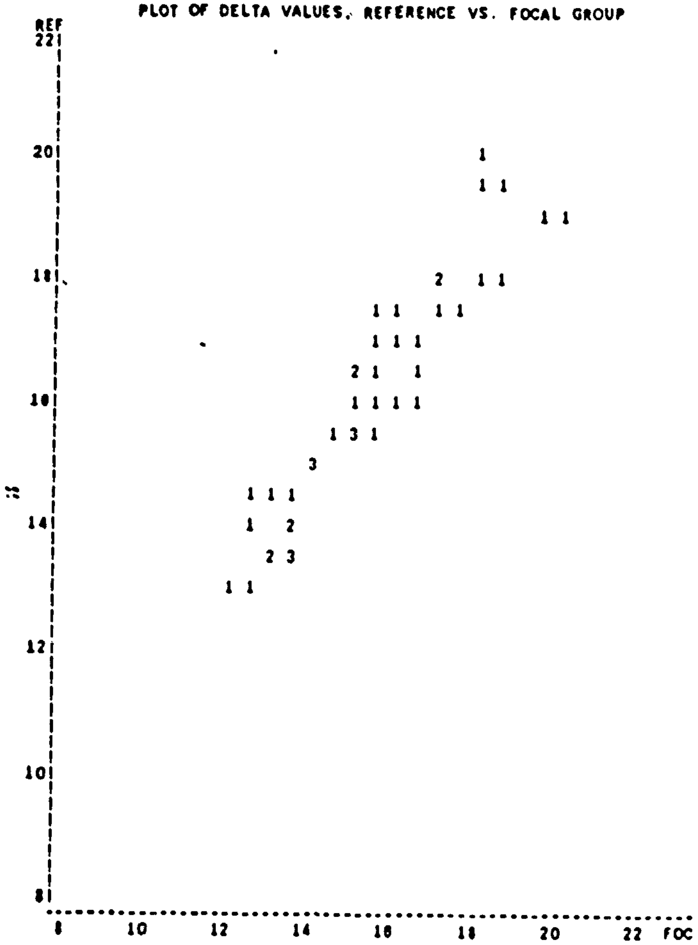

TABLE 5

Item Statistics and DIF Indices for the Delta Plot Method: White VS Black. (N = 500)

NUMBER	P-VALUE REF	P-VALUE FOC	Z-VALUE REF	Z-VALUE FOC	DELTA REF	DELTA FOC	BIAS*
1	.95	.94	1.65	1.55	19.60	19.20	.49
2	.58	.54	.20	.10	13.80	13.40	.70
3	.90	.84	1.28	.99	18.12	16.96	.02
4	.64	.52	.36	.05	14.44	13.20	.10
5	.56	.49	.15	-.03	13.60	12.88	.48
6	.82	.69	.92	.50	16.68	15.00	-.29
7	.94	.91	1.55	1.34	19.20	18.36	.20
8	.60	.48	.25	-.05	14.00	12.80	.14
9	.65	.51	.39	.03	14.56	13.12	-.05
10	.82	.76	.92	.71	16.68	15.84	.29
11	.57	.45	.18	-.13	13.72	12.48	.12
12	.78	.76	.77	.71	16.08	15.84	.73
13	.74	.59	.64	.23	15.56	13.92	-.22
14	.59	.45	.23	-.13	13.92	12.48	-.02
15	.60	.44	.25	-.15	14.00	12.40	-.14
16	.88	.76	1.17	.71	17.68	15.84	-.43
17	.84	.73	.99	.61	16.96	15.44	-.19
18	.77	.64	.74	.36	15.96	14.44	-.15
19	.55	.40	.13	-.25	13.52	12.00	-.06
20	.66	.45	.41	-.13	14.64	12.48	-.55
21	.83	.80	.95	.84	16.80	16.36	.56
22	.89	.80	1.23	.84	17.92	16.36	-.25
23	.89	.87	1.23	1.13	17.92	17.52	.55
24	.96	.96	1.75	1.75	20.00	20.00	.75
25	.89	.79	1.23	.81	17.92	16.24	-.33
26	.84	.66	99	.41	16.96	14.64	-.74
27	.94	.86	1.55	1.08	19.20	17.32	-.52
28	.96	.90	1.75	1.28	20.00	18.12	-.55
29	.87	.78	1.13	.77	17.52	16.08	-.15
30	.84	.72	.99	.58	16.96	15.32	-.27
31	.94	.89	1.55	1.23	19.20	17.92	-.10
32	.84	.72	.99	.58	16.96	15.32	-.27
33	.93	.87	1.48	1.13	18.92	17.52	-.18
34	.68	.54	.47	.10	14.88	13.40	-.09
35	.78	.68	.77	.47	16.08	14.88	.06
36	.85	.76	1.04	.71	17.16	15.84	-.06
37	.71	.60	.55	.25	15.20	14.00	.10
38	.83	.76	.95	.71	16.80	15.84	.20
39	.80	.69	.84	.50	16.36	15.00	-.06
40	.73	.59	.61	.23	15.44	13.92	-.13
41	.71	.59	.5	.23	15.20	13.92	.04
42	.82	.69	.92	.50	16.68	15.00	-.29
43	.67	.54	.44	.10	14.76	13.40	.00
44	.68	.64	.47	.36	14.88	14.44	.63
45	.86	.77	1.08	.74	17.32	15.96	-.09

FIGURE 24 Plot of Delta Values for White and Black Samples. (N=500)

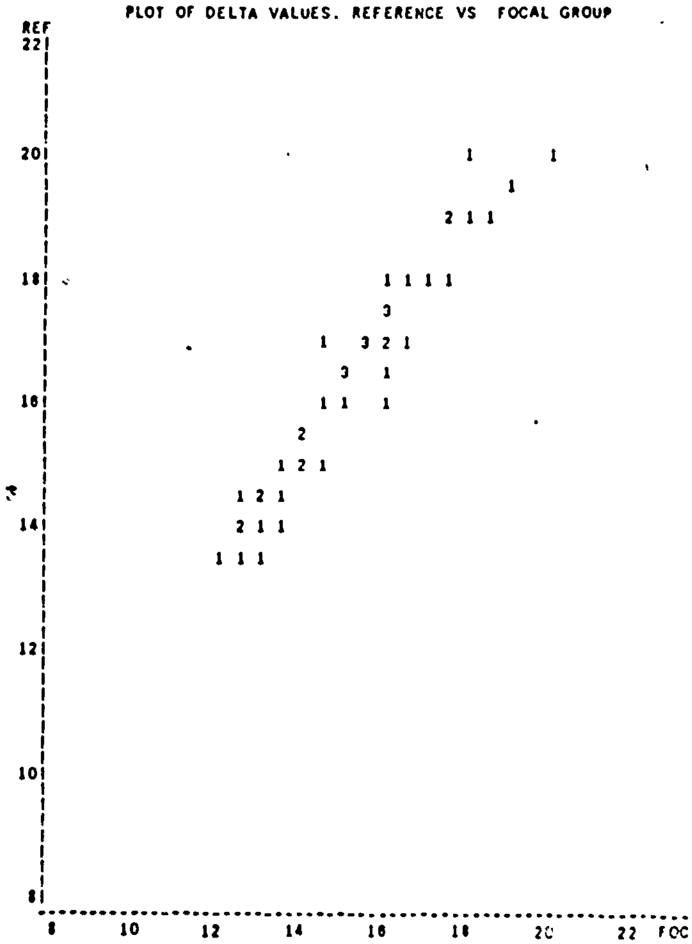

TABLE 6

Item Statistics and DIF Indices for the Delta Plot Method: White VS Black. (N = 750)

NUMBER	P-VALUE REF	P-VALUE FOC	Z-VALUE REF	Z-VALUE FOC	DELTA REF	DELTA FOC	BIAS
1	.96	.91	1.75	1.34	20.00	18.36	.05
2	.56	.52	.15	.05	13.60	13.20	.42
3	.90	.81	1.28	.88	18.12	16.52	-.09
4	.61	.54	.26	.10	14.12	13.40	.22
5	.58	.51	.20	.03	13.80	13.12	.22
6	.83	.69	.95	.50	16.80	15.00	-.35
7	.96	.89	1.75	1.23	20.00	17.92	-.28
8	.60	.49	.25	-.03	14.00	12.88	-.09
9	.67	.52	.44	.05	14.76	13.20	-.35
10	.81	.75	.88	.67	16.52	15.68	.34
11	.51	.47	.03	-.08	13.12	12.68	.34
12	.78	.76	.77	.71	16.08	15.84	.75
13	.73	.61	.61	.28	15.44	14.12	-.11
14	.58	.49	.20	-.03	13.80	12.88	.04
15	.58	.50	.20	.00	13.80	13.00	.13
16	.89	.81	1.23	.88	17.92	16.52	.05
17	.83	.71	.95	.55	16.80	15.20	-.20
18	.79	.60	.81	.25	16.24	14.00	-.73
19	.52	.43	.05	-.18	13.20	12.28	-.01
20	.64	.49	.36	-.03	14.44	12.88	-.38
21	.86	.77	1.08	.74	17.32	15.96	.02
22	.92	.78	1.41	.77	18.64	16.08	-.76
23	.88	.86	1.17	1.08	17.68	17.32	.81
24	.97	.95	1.88	1.65	20.52	19.60	.64
25	.90	.79	1.28	.81	18.12	16.24	-.30
26	.83	.70	.95	.52	16.80	15.08	-.29
27	.94	.87	1.55	1.13	19.20	17.52	-.05
28	.97	.90	1.88	1.28	20.52	18.12	-.47
29	.86	.80	1.08	.84	17.32	16.36	.32
30	.86	.73	1.08	.61	17.32	15.44	-.37
31	.95	.90	1.65	1.28	19.60	18.12	.14
32	.83	.76	.95	.71	16.80	15.84	.28
33	.93	.88	1.48	1.17	18.92	17.68	.26
34	.68	.58	.47	.20	14.88	13.80	.02
35	.79	.67	.81	.44	16.24	14.76	-.16
36	.87	.75	1.13	.67	17.52	15.68	-.32
37	.69	.60	.50	.25	15.00	14.00	.09
38	.84	.72	.99	.58	16.96	15.32	-.22
39	.79	.68	.81	.47	16.24	14.88	-.07
40	.71	.60	.55	.25	15.20	14.00	-.04
41	.67	.57	.44	.18	14.76	13.72	.04
42	.82	.69	.92	.50	16.68	15.00	-.27
43	.66	.54	.41	.10	14.64	13.40	-.12
44	.70	.63	.52	.33	15.08	14.32	.28
45	.84	.80	.99	.84	16.96	16.36	.56

FIGURE 25 Plot of Delta Values for White and Black Samples. (N=750)

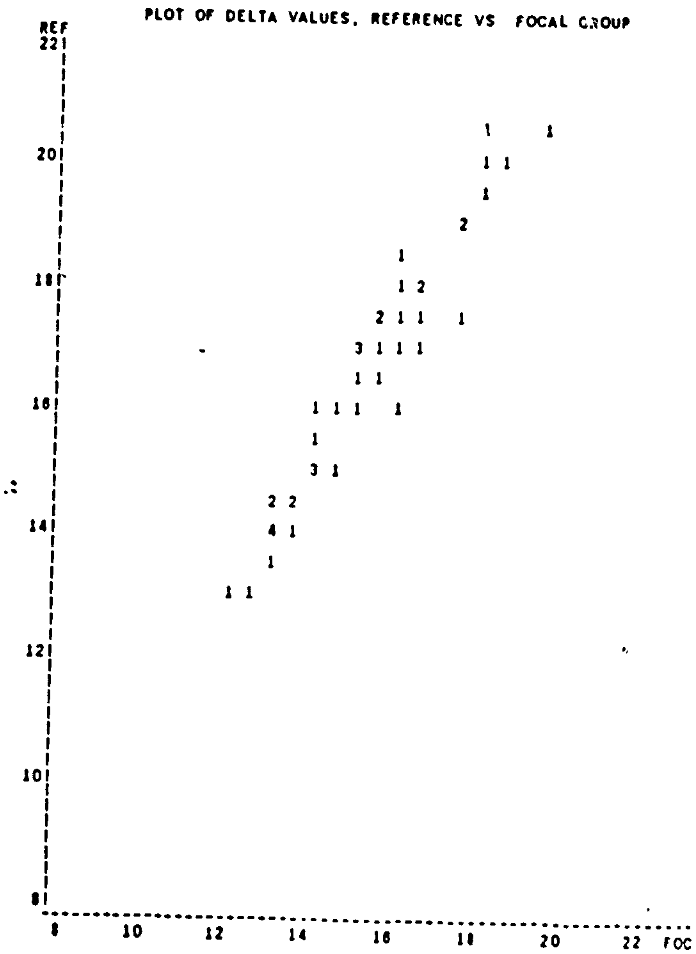

TABLE 7

Item Statistics and DIF Indices for the Delta Plot Method: White VS Black. (N = 1000)

NUMBER	P-VALUE REF	P-VALUE FOC	Z-VALUE REF	Z-VALUE FOC	DELTA REF	DELTA FOC	BIAS*
1	.95	.92	1.65	1.41	19.60	18.64	.30
2	.57	.52	.18	.05	13.72	13.20	.44
3	.89	.81	1.23	.88	17.92	16.52	-.07
4	.61	.53	.28	.08	14.12	13.32	.25
5	.56	.51	.15	.03	13.60	13.12	.47
6	.82	.69	.92	.50	16.68	15.00	-.31
7	.95	.90	1.65	1.28	19.60	18.12	-.07
8	.59	.48	.23	-.05	13.92	12.80	.01
9	.67	.53	.44	.08	14.76	13.32	-.19
10	.80	.74	.84	.64	16.36	15.56	.32
11	.54	.46	.10	-.10	13.40	12.60	.23
12	.79	.76	.81	.71	16.24	15.84	.60
13	.73	.61	.61	.28	15.44	14.12	-.08
14	.60	.42	.25	-.20	14.00	12 20	-.47
15	.58	.50	.20	.00	13.80	13.00	.24
16	.89	.78	1.23	.77	17.92	16.08	-.38
17	.84	.71	.99	.55	16.96	15.20	-.36
18	.78	.63	.77	.33	16.08	14.32	-.38
19	.54	.40	.10	-.25	13.40	12.00	-.20
20	.63	.47	.33	-.08	14.32	12.68	-.35
21	.84	.78	.99	.77	16.96	16.08	.28
22	.91	.81	1.34	.88	18.36	16.52	-.37
23	.89	.87	1.23	1.13	17.92	17.52	.66
24	.98	.95	2.05	1.65	21.20	19.60	-.11
25	.90	.78	1.28	.77	18.12	16.08	-.52
26	.83	.69	.95	.50	16.80	15.00	-.39
27	.94	.88	1.55	1.17	19.20	17.68	-.11
28	.96	.92	1.75	1.41	20.00	18.64	.03
29	.88	.79	1.17	.81	17.68	16.24	-.10
30	.86	.75	1.08	.67	17.32	15.68	-.26
31	.94	.91	1.55	1.34	19.20	18.36	.38
32	.83	.75	.95	.67	16.80	15.68	.10
33	.94	.89	1.55	1.23	19.20	17.92	.06
34	.66	.58	.41	.20	14.64	13.80	.24
35	.78	.67	.77	.44	16.08	14.76	-.06
36	.85	.77	1.04	.74	17.16	15.96	.06
37	.71	.62	.55	.30	15.20	14.20	.14
38	.84	.74	.99	.64	16.96	15.56	-.10
39	.79	.69	.81	.50	16.24	15.00	-.00
40	.73	.58	.61	.20	15.44	13.80	-.32
41	.70	.59	.52	.23	15.08	13.92	.02
42	.81	.71	.88	.55	16.52	15.20	-.05
43	.67	.56	.44	.15	14.76	13.60	.01
44	.70	.61	.52	.23	15.08	14.12	.16
45	.85	.79	1.04	.81	17.16	16.24	.26

FIGURE 26 Plot of Delta Values for White and Black Samples. (N=1000)

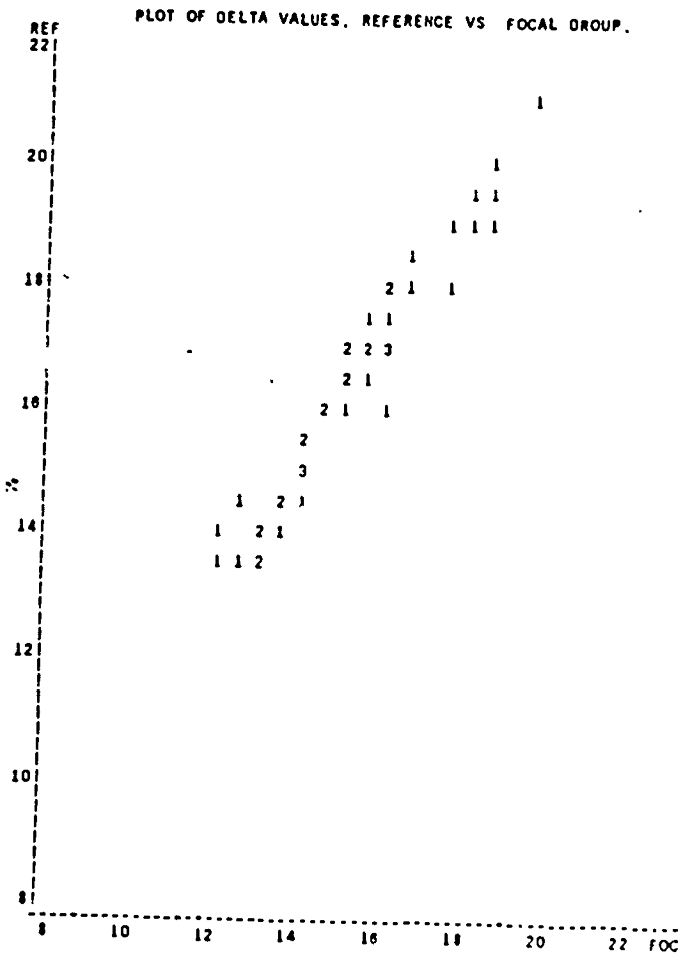

TABLE 8

Item Statistics and DIF Indices for the Delta Plot Method: Male VS Female. (N = 200)

NUMBER	P-VALUE REF	P-VALUE FOC	Z-VALUE REF	Z-VALUE FOC	DELTA REF	DELTA FOC	BIAS*
1	.92	.93	1.41	1.48	18.64	18.92	-.66
2	.60	.51	.25	.03	14.00	13.12	-.03
3	.84	.87	.99	1.13	16.96	17.52	-.04
4	.58	.58	.20	.20	13.80	13.80	.52
5	.54	.53	.10	.08	13.40	13.32	.58
6	.79	.75	.81	.67	16.24	15.68	-.46
7	.94	.91	1.55	1.34	19.20	18.36	-1.44
8	.60	.53	.25	.08	14.00	13.32	.09
9	.66	.54	.41	.10	14.64	13.40	-.40
10	.77	.76	.74	.71	15.96	15.84	-.14
11	.51	.50	.03	.00	13.12	13.00	.64
12	.77	.79	.74	.81	15.96	16.24	.08
13	.65	.66	.39	.41	14.56	14.64	.35
14	.56	.53	.15	.08	13.60	13.32	.42
15	.60	.54	.25	.10	14.00	13.40	.13
16	.83	.87	.95	1.13	16.80	17.52	.09
17	.81	.77	.88	.74	16.52	15.96	-.54
18	.69	.78	.50	.77	15.00	16.08	.79
19	.55	.42	.13	-.20	13.52	12.20	-.14
20	.57	.57	.18	.18	13.72	13.72	.54
21	.80	.80	.84	.84	16.36	16.36	-.19
22	.85	.89	1.04	1.23	17.16	17.92	.02
23	.85	.91	1.04	1.34	17.16	18.36	.26
24	.96	1.00	1.75	3.00	20.00	25.00	1.59
25	.84	.84	.99	.99	16.96	16.96	-.35
26	.77	.72	.74	.58	15.96	15.32	-.43
27	.87	.95	1.13	1.65	17.52	19.60	.65
28	.90	.96	1.28	1.75	18.12	20.00	.37
29	.82	.88	.92	1.17	16.68	17.68	.28
30	.82	.80	.92	.84	16.68	16.36	-.45
31	.91	.93	1.34	1.48	18.36	18.92	-.43
32	.80	.79	.84	.81	16.36	16.24	-.25
33	.88	.90	1.17	1.28	17.68	18.12	-.31
34	.61	.57	.28	.18	14.12	13.72	.21
35	.74	.71	.64	.55	15.56	15.20	-.17
36	.81	.80	.88	.84	16.52	16.36	-.32
37	.69	.61	.50	.28	15.00	14.12	-.30
38	.80	.76	.84	.71	16.36	15.84	-.48
39	.72	.73	.58	.61	15.32	15.44	.17
40	.69	.60	.50	.25	15.00	14.00	-.37
41	.65	.59	.39	.23	14.56	13.92	-.05
42	.78	.77	.77	.74	16.08	15.96	-.18
43	.63	.60	.33	.25	14.32	14.00	.20
44	.70	.73	.52	.61	15.08	15.44	.37
45	.83	.84	.95	.99	16.80	16.96	-.22

FIGURE 27 Plot of Delta Values for Male and Female Samples. (N=200)

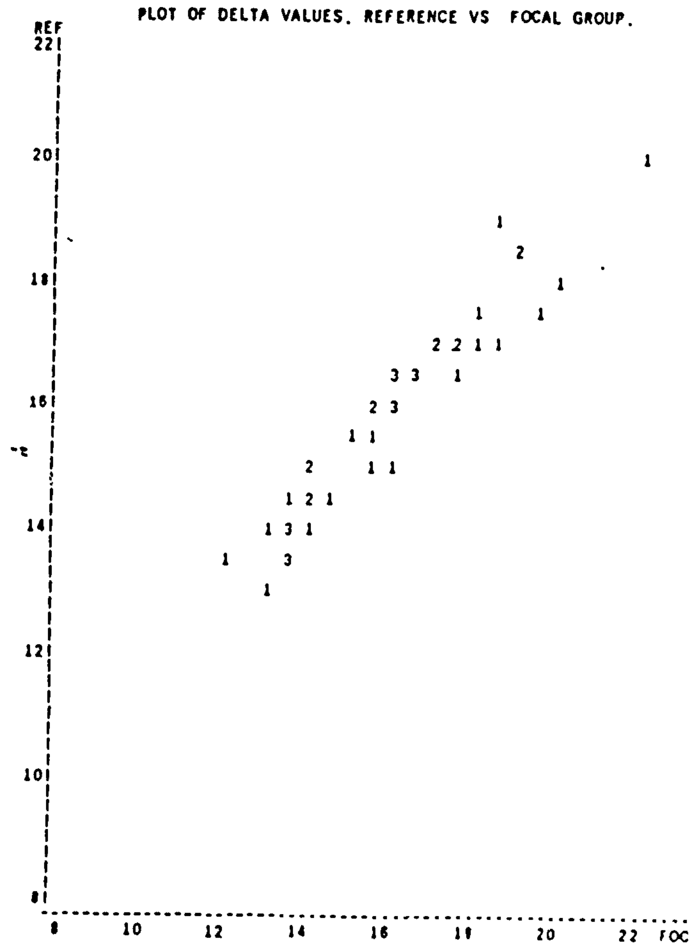

TABLE 9

Item Statistics and DIF Indices for the Delta Plot Method: Male VS Female. (N = 500)

NUMBER	P-VALUE REF	P-VALUE FOC	Z-VALUE REF	Z-VALUE FOC	DELTA REF	DELTA FOC	BIAS*
1	.94	.92	1.55	1.41	19.20	18.64	-.40
2	.59	.51	.23	.03	13.92	13.12	-.54
3	.83	.84	.95	.99	16.80	16.96	.12
4	.59	.58	.23	.20	13.92	13.80	-.06
5	.55	.51	.13	.03	13.52	13.12	-.25
6	.77	.73	.74	.61	15.96	15.44	-.35
7	.94	.90	1.55	1.28	19.20	18.12	-.77
8	.57	.53	.18	.08	13.72	13.32	-.25
9	.64	.54	.36	.10	14.44	13.40	-.71
10	.75	.76	.67	.71	15.68	15.84	.13
11	.50	.52	.00	.05	13.00	13.20	.17
12	.72	.76	.58	.71	15.32	15.84	.38
13	.67	.66	.44	.41	14.76	14.64	-.06
14	.54	.52	.10	.05	13.40	13.20	-.11
15	.56	.52	.15	.05	13.60	13.20	-.25
16	.84	.79	.99	.81	16.96	16.24	-.50
17	.76	.78	.71	.77	15.84	16.08	.18
18	.69	.73	.50	.61	15.00	15.44	.33
19	.47	.49	-.08	-.03	12.68	12.88	.18
20	.58	.58	.20	.20	13.80	13.80	.03
21	.80	.82	.84	.92	16.36	16.68	.24
22	.82	.83	.92	.95	16.68	16.80	.09
23	.83	.86	.95	1.08	16.80	17.32	.38
24	.94	.93	1.55	1.48	19.20	18.92	-.20
25	.84	.83	.99	.95	16.96	16.80	-.10
26	.78	.72	.77	.58	16.08	15.32	-.52
27	.88	.92	1.17	1.41	17.68	18.64	.68
28	.91	.91	1.34	1.34	13.36	18.36	-.00
29	.79	.82	.81	.92	16.24	16.68	.32
30	.77	.78	.74	.77	15.96	16.08	.10
31	.91	.91	1.34	1.34	18.36	18.36	-.00
32	.78	.78	.77	.77	16.08	16.08	.01
33	.90	.88	1.28	1.17	18.12	17.68	-.31
34	.62	.61	.30	.28	14.20	14.12	-.03
35	.71	.74	.55	.64	15.20	15.56	.27
36	.77	.80	.74	.84	15.96	16.36	.30
37	.65	.71	.39	.55	14.56	15.20	.47
38	.77	.76	.74	.71	15.96	15.84	-.07
39	.71	.77	.55	.74	15.20	15.96	.55
40	.62	.68	.30	.47	14.20	14.88	.50
41	.61	.64	.28	.36	14.12	14.44	.25
42	.79	.73	.81	.61	16.24	15.44	-.55
43	.63	.62	.33	.30	14.32	14.20	-.06
44	.66	.65	.41	.39	14.64	14.56	-.03
45	.80	.84	.84	.99	16.36	16.96	.43

FIGURE 28 Plot of Delta Values for Male and Female Samples. (N=500)

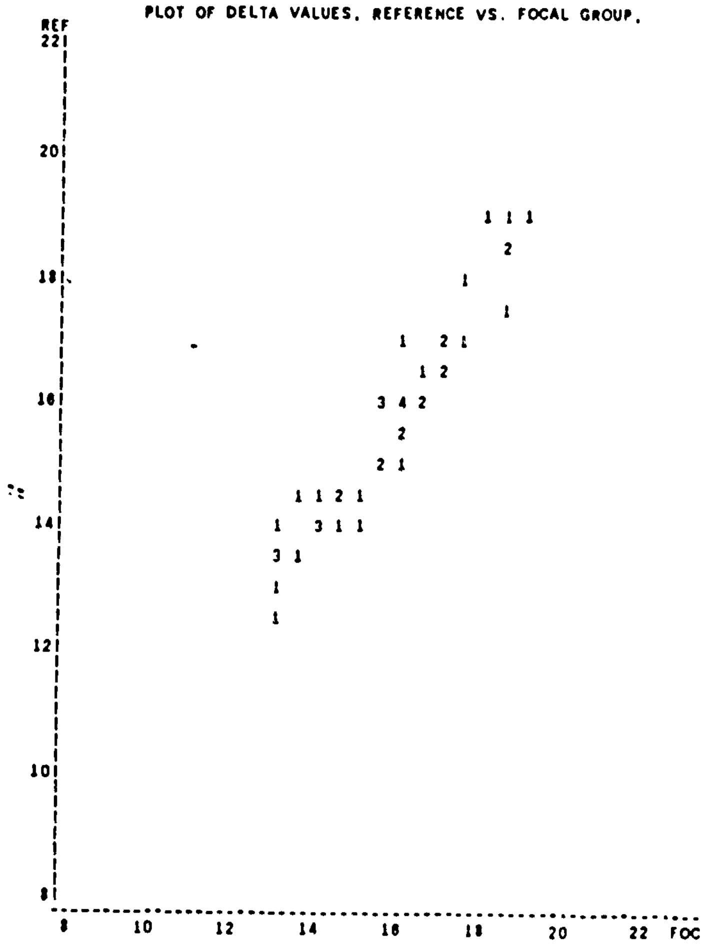

TABLE 10

Item Statistics and DIF Indices for the Delta Plot Method: Male VS Female. (N = 750)

NUMBER	P-VALUE REF	P-VALUE FOC	Z-VALUE REF	Z-VALUE FOC	DELTA REF	DELTA FOC	BIAS'
1	.93	.89	1.48	1.23	18.92	17.92	-.60
2	.59	.52	.23	.05	13.92	13.20	-.23
3	.84	.84	.99	.99	16.96	16.96	.16
4	.59	.59	.23	.23	13.92	13.92	.26
5	.56	.53	.15	.08	13.60	13.32	.08
6	.78	.72	.77	.58	16.08	15.32	-.34
7	.91	.88	1.34	1.17	18.36	17.68	-.36
8	.55	.47	.13	-.08	13.52	12.68	-.30
9	.67	.56	.44	.15	14.76	13.60	-.57
10	.77	.76	.74	.71	15.96	15.84	.11
11	.50	.52	.00	.05	13.00	13.20	.43
12	.76	.74	.71	.64	15.84	15.56	.00
13	.71	.67	.55	.44	15.20	14.76	-.09
14	.53	.49	.08	-.03	13.32	12.88	-.02
15	.56	.49	.15	-.03	13.60	12.88	-.22
16	.83	.80	.95	.84	16.80	16.36	-.14
17	.78	.77	.77	.74	16.08	15.96	.10
18	.69	.74	.50	.64	15.00	15.56	.61
19	.49	.44	-.03	-.15	12.88	12.40	-.03
20	.58	.50	.20	.00	13.80	13.00	-.28
21	.80	.80	.84	.84	16.36	16.36	.18
22	.83	.84	.95	.99	16.80	16.96	.27
23	.84	.86	.99	1.08	16.96	17.32	.40
24	.92	.93	1.41	1.48	18.64	18.92	.29
25	.84	.81	.99	.88	16.96	16.52	-.15
26	.78	.70	.77	.52	16.08	15.08	-.50
27	.87	.89	1.13	1.23	17.52	17.92	.41
28	.90	.89	1.28	1.23	18.12	17.92	-.02
29	.82	.81	.92	.88	16.68	16.52	.05
30	.80	.78	.84	.77	16.36	16.08	-.02
31	.89	.87	1.23	1.13	17.92	17.52	-.15
32	.79	.77	.81	.74	16.24	15.96	-.01
33	.88	.85	1.17	1.04	17.68	17.16	-.23
34	.62	.58	.30	.20	14.20	13.80	-.02
35	.73	.73	.61	.61	15.44	15.44	.21
36	.81	.78	.88	.77	16.52	16.08	-.13
37	.67	.64	.44	.36	14.76	14.44	.01
38	.78	.75	.77	.67	16.08	15.68	-.09
39	.73	.73	.61	.61	15.44	15.44	.21
40	.65	.67	.39	.44	14.56	14.76	.38
41	.65	.64	.39	.36	14.56	14.44	.16
42	.78	.73	.77	.61	16.08	15.44	-.25
43	.63	.60	.33	.25	14.32	14.00	.03
44	.64	.65	.36	.39	14.44	14.56	.33
45	.80	.79	.84	.81	16.36	16.24	.09

FIGURE 29 Plot of Delta Values for Male and Female Samples. (N=750)

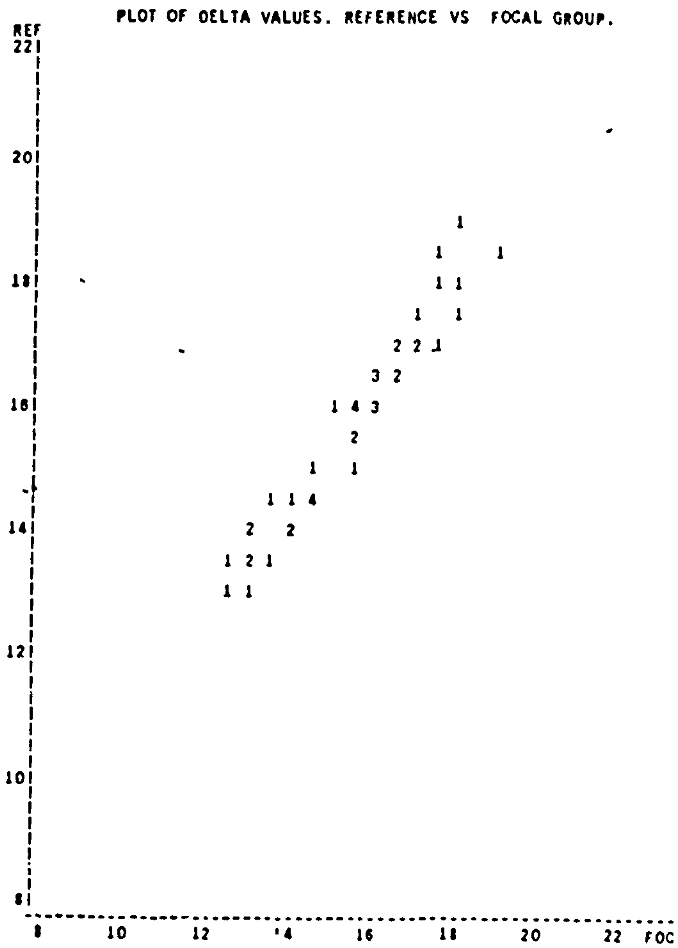

TABLE 11

Item Statistics and DIF Indices for the Delta Plot Method: Male VS Female. (N = 1000)

NUMBER	P-VALUE REF	P-VALUE FOC	Z-VALUE REF	Z-VALUE FOC	DELTA REF	DELTA FOC	BIAS
1	.92	.89	1.41	1.23	18.64	17.92	-.48
2	.59	.51	.23	.03	13.92	13.12	-.38
3	.84	.85	.99	1 04	16.96	17.16	.21
4	.58	.57	.20	.18	13.80	13.72	.12
5	.55	.54	.13	.10	13.52	13.40	.10
6	.76	.74	.71	.64	15.84	15.56	-.08
7	.92	.89	1.41	1.23	18.64	17.92	-.48
8	.57	.53	.18	.08	13.72	13.32	-.10
9	.65	.55	.39	.13	14.56	13.52	-.57
10	.78	.76	.77	.71	16.08	15.84	-.06
11	.50	.50	.00	.00	13.00	13.00	.20
12	.75	.75	.67	.67	15.68	15.68	.11
13	.68	.65	.47	.39	14.88	14.56	-.08
14	.54	.52	.10	.05	13.40	13.20	.05
15	.58	.49	.20	-.03	13.80	12.88	-.46
16	.83	.80	.95	.84	16.80	16.36	-.22
17	.79	.76	.81	.71	16.24	15.84	-.18
18	.70	.74	.52	.64	15.08	15.56	.47
19	.49	.46	-.03	-.10	12.88	12.60	.01
20	.58	.53	.20	.08	13.80	13.32	-.16
21	.79	.79	.81	.81	16.24	16.24	.10
22	.82	.84	.92	.99	16.68	16.96	.28
23	.84	.87	.99	1.13	16.96	17.52	.46
24	.92	.93	1.41	1.48	18.64	18.92	.21
25	.85	.82	1.04	.92	17.16	16.68	-.26
26	.78	.74	.77	.64	16.08	15.56	-.26
27	.88	.90	1.17	1.28	17.68	18.12	.36
28	.91	.91	1.34	1.34	18.36	18.36	.03
29	.82	.81	.92	.88	16.68	16.52	-.03
30	.78	.78	.77	.77	16.08	16.08	.10
31	.90	.88	1.28	1.17	18.12	17.68	-.27
32	.79	.77	.81	.74	16.24	15.96	-.10
33	.88	.87	1.17	1.13	17.68	17.52	-.06
34	.62	.58	.30	.20	14.20	13.80	-.12
35	.70	.76	.52	.71	15.08	15.84	.66
36	.79	.76	.81	.71	16.24	15.84	-.18
37	.68	.65	.47	.39	14.88	14.56	-.08
38	.78	.75	.77	.67	16.08	15.68	-.17
39	.73	.74	.61	.64	15.44	15.56	.20
40	.64	.64	.36	.36	14.44	14.44	.15
41	.61	.63	.28	.33	14.12	14.32	.30
42	.76	.73	.71	.61	15.84	15.44	-.17
43	.60	.61	.25	.28	14.00	14.12	.25
44	.64	.67	.36	.44	14.44	14.76	.37
45	.80	.81	.84	.88	16.36	16.52	.20

FIGURE 30 Plot of Delta Values for Male and Female Samples. (N=1000)

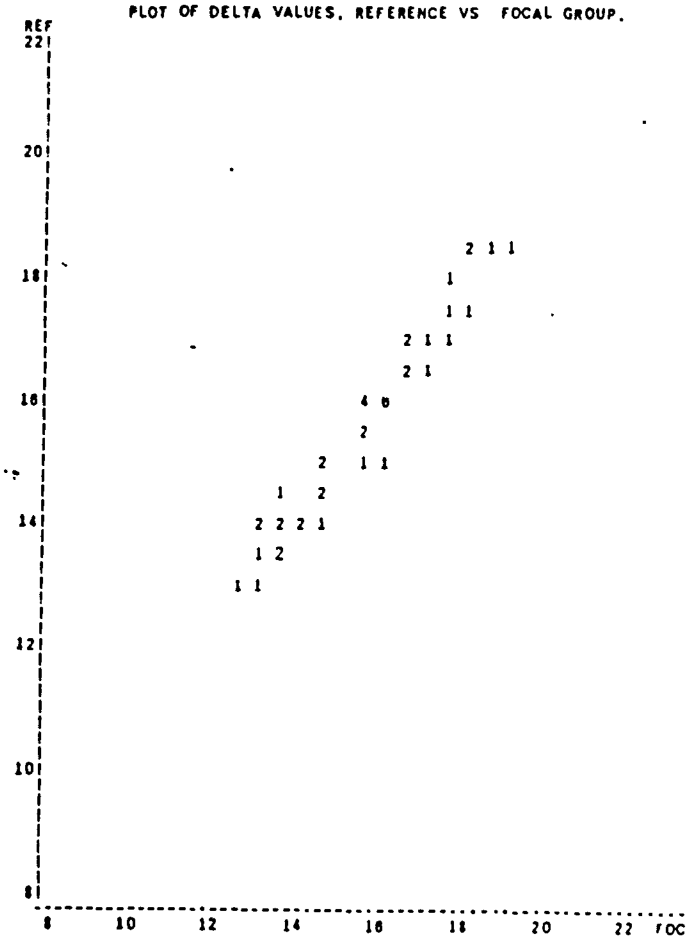

TABLE 12

Stability of DIF Statistics Across Independent Samples (Black vs White)

	N=200	N=500	N=750	N=1000
N=200				
Delta	------	0.426	0.137	0.159
Rasch	------	0.148	0.158	0.242
M-H	------	0.303	0.352	0.311
N=500				
Delta		------	0.416	0.412
Rasch		------	0.427	0.498
M-H		------	0.468	0.719
N=750				
Delta			------	0.354
Rasch			------	0.486
M-H			------	0.738

Note: Stability indicated by rank order correlations.

TABLE 13

Stability of DIF Statistics Across Independent Samples (Male vs Female)

	N=200	N=500	N=750	N=1000
N=200				
Delta	------	0.112	0.116	0.104
Rasch	------	0.169	0.481	0.309
M-H	------	0.313	0.358	0.401
N=500				
Delta		------	0.425	0.317
Rasch		------	0.335	0.468
M-H		------	0.244	0.345
N=750				
Delta			------	0.602
Rasch			------	0.682
M-H			------	0.518

Note: Stability indicated by rank order correlations.

TABLE 14

Stability of DIF Statistics Across Independent Samples (Black vs White)

	750 vs 1000		500 vs 1000		200 vs 1000	
	P1	P2	P1	P2	P1	P2
Delta	0.950	0.04 (2)	0.950	0.02 (1)	1.000	0.0 (0)
Rasch	0.930	0.04 (2)	0.930	0.02 (1)	0.930	0.02 (1)
M-H	0.840	0.75 (34)	0.880	0.73 (33)	0.220	0.06 (3)

Note: P1=proportion of total hits; P2=proportion of true positives; Numbers in parenthesis indicate the number of items.

TABLE 15

Stability of DIF Statistics Across Independent Samples (Male vs Female)

	750 vs 1000		500 vs 1000		200 vs 1000	
	P1	P2	P1	P2	P1	P2
Delta	1.000	0.00 (0)	1.000	0.00 (0)	1.000	0.00 (0)
Rasch	0.970	0.00 (0)	0.970	0.00 (0)	0.930	0.02 (1)
M-H	0.840	0.04 (2)	0.800	0.02 (1)	1.000	0.00 (0)

Note: P1=proportion of total hits; P2=proportion of true positives; Numbers in parenthesis indicate the number of items.

TABLE 16

Agreement of DIF Statistics Across Techniques. (Black vs White)

	Delta	Rasch	M-H	3 Parameter
Delta				
N=200	-------	0.206	0.006	0.012
N=500	-------	0.906	0.195	0.086
N=750	-------	0.886	0.024	0.113
N=1000	-------	0.901	0.015	0.476
Rasch				
N=200		------	0.086	0.126
N=500		------	0.108	0.350
N=750		------	0.094	0.410
N=1000		------	0.033	0.535
M-H				
N=200			------	0.094
N=500			------	0.367
N=750			------	0.003
N=1000			------	0.236

Note: Agreement indicated by Rank Order Correlations.

TABLE 17

Agreement of DIF Statistics Across Techniques. (Male vs Female)

	Delta	Rasch	M-H	3 Parameter
Delta				
N=200	-------	0.072	0.119	0.065
N=500	-------	0.867	0.184	0.136
N=750	-------	0.901	0.213	0.051
N=1000	-------	0.880	0.062	0.218
Rasch				
N=200		------	0.265	0.212
N=500		------	0.212	0.310
N=750		------	0.208	0.371
N=1000		------	0.310	0.510
M-H				
N=200			------	0.033
N=500			------	0.269
N=750			------	0.244
N=1000			------	0.258

Note: Agreement indicated by Rank Order Correlations.

TABLE 18

Agreement of Three DIF Techniques with the Three-Parameter Model
(Black vs. White)

	N = 1000		N = 750		N = 500		N = 200	
	P1	P2	P1	P2	P1	P2	P1	P2
Delta	0.930	0.02 (1)	0.880	0.02 (1)	0.930	0.02 (1)	0.930	0.000
Rasch	0.890	0.02 (1)	0.910	0.02 (1)	0.910	0.00 (0)	0.910	0.000
M-H	0.220	0.06 (3)	0.240	0.04 (2)	0.260	0.04 (2)	0.870	0.000

Note. P1=proportion of total hits; P2=proportion of true positive; numbers in parentheses indicate the number of items.

TABLE 19

Agreement Across Three-DIF Techniques with the Three-Parameter Model
(Male vs. Female)

	N = 1000		N = 750		N = 500		N = 200	
	P1	P2	P1	P2	P1	P2	P1	P2
Delta	0.970	0.02 (1)	0.960	0.00 (0)	0.960	0.00 (0)	0.960	0.000
Rasch	0.970	0.00 (0)	0.950	0.00 (0)	0.950	0.00 (0)	0.950	0.000
M-H	0.880	0.04 (2)	0.950	0.02 (1)	0.880	0.00 (0)	0.950	0.000

Note. P1=proportion of total hits; P2=proportion of true positive; Numbers in parantheses indicate the number of items.

END

U.S. Dept. of Education

Office of Education
Research and
Improvement (OERI)

ERIC

Date Filmed

March 21,1991